December 2020

Examining the Notion of the Boundary Object in Information Systems: the Transdisciplinary Oeuvre of Cognitive Science

Laura Elien Ridenour
University of Wisconsin-Milwaukee

EXAMINING THE NOTION OF THE BOUNDARY OBJECT IN INFORMATION SYSTEMS: THE TRANSDISCIPLINARY OEUVRE OF COGNITIVE SCIENCE

by

Laura Ridenour

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Information Studies

at

University of Wisconsin-Milwaukee

December 2020

ABSTRACT
EXAMINING THE NOTION OF THE BOUNDARY OBJECT IN INFORMATION SYSTEMS: THE TRANSDISCIPLINARY OEUVRE OF COGNITIVE SCIENCE

by

Laura Ridenour

The University of Wisconsin-Milwaukee, 2020
Under the Supervision of Professor Richard P. Smiraglia

This study examined the transdisciplinary area of cognitive science, and was framed around the sociological notion of the boundary object. Harmonizing theoretical and technical approaches, methods introduced in this work moved beyond qualitative study practices traditional to boundary object theory work to a mixed-methods data-driven approach. Bibliometric Web of Science data, enriched with National Science Foundation (NSF) journal classifications, formed the foundation from which a seed-and-expand dataset were created from journals containing the string cogni* and their cited articles for the years 2006-2016. This two-tiered dataset allowed for the analysis of boundary-spanning interdisciplinary concepts, as identified by noun phrases, and their inhabitance within the intellectual space of the NSF taxonomy. The most interdisciplinary concepts were analyzed for their conceptual periphera using term co-occurrences, and the underlying sociological structures of co-authorship. Two concepts met the criterion of publication in all six core-level NSF disciplines resulted in two for this analysis: "children's," and "case study." Clearer clusters of term co-occurrences were present for "children's" than were for "case study," demonstrating the conceptual periphera. The underlying social structures for "children's" were more interconnected than those for "case study." The findings of this study suggest that different types of research problems, in conjunction with the methodology used to explore them, may be more useful to pinpoint boundary-inhabiting interdisciplinary epistemologies and other conceptual phenomena than the examination of broadly defined boundary-spanning concepts alone.

TABLE OF CONTENTS

List of Figures **viii**

List of Tables **ix**

List of Abbreviations **x**

Chapter 1 - Introduction **1**

1 Introduction **1**

1.1 Background 4

1.2 Boundaries 6

1.3 Concepts 8

1.4 Classification: From concepts to Shared Understandings 11

1.5 Domains, disciplines, and boundary crossing 11

1.6 Interdisciplinarity 16

1.7 Concept Translation 18

1.8 Conclusion 19

1.8.1 Research Questions 20

1.8.2 Outline of the Dissertation 20

Chapter 2 - Literature Review **23**

2 Literature Review **23**

2.1 Research on Boundary Objects 23

2.2 Disciplines that study and discuss boundary objects 25

2.2.1 Sociology 25

2.2.2 Social Network Analysis 26

2.2.3 Medicine 26

2.2.4 Education and Psychology 27

2.2.5 Management, Organizational Science, and Information Systems . 29

2.2.6 Information Science 32

2.3 Boundary Objects and Interdisciplinarity 34
2.4 Measures of Interdisciplinarity 36
2.4.1 Classification and Indexing: Respecting Epistemological Stances . 37
2.4.2 Measuring Indexing Terms: Exhaustivity and Specificity 38
2.4.3 Classification of Journals 39
2.5 Conclusion 40

Chapter 3 - Review of Methods **42**

3 Review of Methods **42**
3.1 Text Processing and Informetrics 42
3.2 Citation Analysis 42
3.3 Mapping Citation Flows 43
3.4 Co-citation Analysis 43
3.5 Bibliographic Coupling 44
3.6 Author Co-citation Analysis 45
3.7 Co-word Analysis 46
3.8 Topic Delineation and Community Detection 49
3.9 Applications of Text Processing to Interdisciplinarity 51
3.10 Metrics-based Approaches to Text Processing and Information Retrieval . 52
3.11 Hybridization of Methods 54
3.12 Support Vector Machines and Bibliometrics 54
3.13 Author Topic Modeling 54
3.14 Citation Analysis for the Examination of Interdisciplinarity 55
3.15 Visualization and Science Mapping 56
3.16 Disciplinary Diversity and Network Analysis 58
3.17 Systems for Addressing Interdisciplinary Classification 59
3.18 Conclusion 60

Chapter 4 - Methods **62**

4 Methods **62**

4.1 Data 63

4.2 Triangulating Boundary Concepts 64

4.3 Analysis of Boundary Concepts 67

4.4 Methodological Pipeline 68

4.5 Focus Restrictions 69

4.6 The Cognitive Science Domain 70

4.7 Examining Boundary Concepts: The Domain of Cognitive Science 70

4.8 Examining Boundary Concepts: The Cognitive Science Core 71

4.9 Analysis of Authors, Interdisciplinarity, and Collaboration 72

4.10 Distinguishing Between Disciplines 74

4.11 Limitations 75

4.12 Selection of Methods 76

4.13 Conclusion 77

Chapter 5 - Results **78**

5 Results **78**

5.1 Epistemological Boundaries 78

5.2 Terminology and Boundaries 82

5.3 Core-level Interdisciplinary Concepts 85

5.4 Results: Entire Domain 88

5.5 Core: Analysis of Concepts 88

5.6 Authorship Analysis 96

5.6.1 Author Boundary Crossing 96

5.6.2 Co-authorship Surrounding Core Boundary Spanning Concepts . 97

5.6.3 Coauthorship: Case Study 101

5.7 Conclusion 104

Chapter 6 - Discussion **105**

6 Discussion **105**

6.1 Assumptions . . . 108

6.2 Web of Science Data . . . 109

6.3 Moving Outside of Prepackaged Tools . . . 110

6.4 Examining Concepts in Natural Language . . . 111

6.5 Concept-bearing Terms . . . 114

6.6 Potential Applications . . . 114

Chapter 7 - Future Work **116**

7 Future Work **116**

7.1 Pivot Points . . . 118

7.2 Classifying Concepts . . . 119

7.3 Collocating Concepts . . . 119

7.4 Representing the Structure of Cognitive Science . . . 120

7.5 Conclusion . . . 121

References **122**

Appendices **139**

7.6 Appendix A: Code . . . 139

7.7 Appendix B: Tables . . . 143

Curriculum Vitae **146**

LIST OF FIGURES

3.1 Co-citation Analysis 44
3.2 Bibliographic Coupling 45
4.1 Interdisciplinary Span of Concepts 68
4.2 Data Selection 69
5.1 Epistemological Spread of Cognitive Science: Discipline Level 80
5.2 Publications by Level by GrDisc by Year 82
5.3 Boundary Concepts: Burst Analysis 84
5.4 Total Number of Core Boundary Concepts per Year 86
5.5 Total Number of Core Boundary Concepts per Year: Fitted Time-Series Model 88
5.6 Core: Case Study Distribution of Disciplines 90
5.7 Core: Children's Distribution of Disciplines 92
5.8 Children's: Co-occurring Concepts 93
5.9 Case Study: Co-occurring Concepts 95
5.10 Children's Co-authorship Network: All Authors 99
5.11 Children's Co-authorship Network: Largest Connected Component 100
5.12 All Case Study Co-authors 102
5.13 Case Study Co-authors: Largest Connected Component 103

LIST OF TABLES

3.1 Rafols and Meyer's Patterns of Interdisciplinary Integration 59

4.1 Research Questions and Methods 64

4.2 High-level Discipline Recoding 64

4.3 Recoded Disciplines 65

4.4 Data Collection Step 65

5.1 Num. Publications by Level, by Gr. Disc, by Disc, by Year 81

5.2 Number of Core Publications Per Year by Category 81

5.3 Number of Boundary Concepts per Year: Core Data 85

5.4 Number of Core Boundary Concepts per Year Occurring 16 or More Times 87

5.5 Number of Publications per Year (Core and Extended) 89

5.6 Article Count by Year by Level 89

5.7 Publishing Authors by Year by Category by Level 97

5.8 Children's Coauthorship: Network Statistics 98

5.9 Case Study: Network Statistics 101

7.1 Concept Concentration: Title and Abstract vs. Title 117

7.2 Level, High-level Discipline, Discipline, and Article Counts 143

7.3 Full Category Names, Recoded Data, Num. Records, and Num. Abstracts 143

7.4 Children's Coauthorship by Cluster Aligned to NSF Taxonomy 144

LIST OF ABBREVIATIONS

BC - Boundary Concept

BO - Boundary Object

CHAT - Cultural-Historical Activity Theory

Disc - Discipline GrDisc - Grand Discipline (High-level discipline)

LDA - Latent Dirichlet Allocation

LSI - Latent Semantic Indexing

HMM - Hidden Markov Model

NSE - Natural Sciences and Engineering

NSF - National Science Foundation

SSH - Social Sciences and Humanities

Spec - Specialty

WoS - Web of Science

For Ramona,

and her Papa

ACKNOWLEDGEMENTS

I would like to thank my committee for agreeing to let me craft an ambitious proof-of-concept. My advisor, Richard Smiraglia, was instrumental in giving me the freedom to pursue a novel and creative approach to examine interdisciplinarity. When I asked him if he would like to be my advisor, his response was "oh, absolutely." Richard also helped me to realize that just because I did not neatly fit within the bounds of what is known as *knowledge organization*, did not mean that I did not belong in that intellectual space. He also provided me with a great deal of professional support, helping me integrate into a scholarly sphere greater than one specialty while also supporting my leadership strengths. I thank Dietmar Wolfram for his careful and thorough consideration while teaching and in providing feedback for not only coursework, but professional submissions. Thanks to Iris Xie for teaching me to express qualitative and complex concepts in tables, and for her support in strategizing how to combine academic work and motherhood. I thank Rick Szostak for his wholehearted support of my "why not?" line of inquiry, and his modeling of how to handle criticism of new ideas in an open forum with grace. Finally, I thank Margaret Kipp; she was instrumental in helping encourage me to troubleshoot fiddly file formats and in making me sit down and actually do it instead of just talking about it.

I am forever indebted to Andrea Scharnhorst. She gladly gifted me not only her time, but shared her in-depth knowledge of science of science, and a wealth of literature that would have taken me many months to curate. Andrea not only happily agreed to be on my committee without even having met me, she introduced me to a circle of metrics-oriented researchers at ASIS&T in Copenhagen. She then invited me to present at the COST Knowescape workshop in Warsaw. At Knowescape, I was introduced to Loet Leydesdorff and Kevin Boyack, whose previous work provided much of the backbone of the metrics-based approaches influencing my work. In a sense, she is responsible for much of the direction of this work. I am grateful to her for the opportunities she provided me, as well as the warmth with which she has treated me and the interest she has shown in my ideas. Unfortunately, she was unable to see my dissertation through to the end for personal reasons.

Next, I would like to thank Stefanie Haustein, who invited me to Montreal to present for the members of Observatoire des sciences et technologies, Centre interuniversitaire de recherche sur la science et la technologie, UQAM, and to discuss my research with Vincent Larivière. Vincent was gracious enough to connect me with the excellent value-added Web of Science data used in this work. I am also grateful to Diane Marie Plante for sharing her deep understanding of interdisciplinarity, as well as what it means to be interdisciplinary. While in Montréal, I was glad to forge now long-standing and productive research partnerships, especially with Maxime Sainte-Marie.

In addition to those who directly influenced this work, I would like to thank the many individuals for helping shape me professionally, including Nadine Kozak, Kathryn La Barre, Barbara Kwašnik, Ron Day, Joe Tennis, Staša Milojević, and Laretta Henderson. I'd also like to thank my Ph.D. peers, some of whom have since graduated and moved on. They include Nick Proferes, Hyoungjoo Park, Yuehua Zhao, Yanyan Wang, Shannon Crawford-Barniskiss, Vanessa Schlais, and Meghan Dowell. I would also like to acknowledge Amy Waldman, Daniel Collette, B Charles, and my daughter Ramona; they all inspired and pushed me to keep going through the darkest hours of this time in my life. I must also express my deep gratitude to Chrissy Boehm and the staff of the Buttercup Room for the excellent care Ramona received while at the Children's Learning Center at the University of Wisconsin-Milwaukee.

I am grateful to the assistance provided to me by countless friends and family, especially those who (sometimes literally) came with us on a weird and wild transcontinental academic journey: Sarah Knudsen; Brigid Rankowski; Katherine Ahnberg; Tiffany Smith; Aurora Keeler; and my mother, Catherine Ridenour. I'm particularly grateful to my mother for letting me borrow my father, John Ridenour, and to my father for gifting Ramona and me with his time, patience, and presence. I am also thankful for the supportive nature of my siblings–Jack and Meg, you'll both go far. I must also thank my at-home team of feline critics, who made formatting my work in LaTeX an adventure. Because the list of those who have touched me, and by extension this work (even through COVID), is so long...I thank my village.

Chapter 1

Introduction

1 Introduction

This study is an examination of an eleven year section of the known transdisciplinary area, cognitive science. The objective was to identify core concepts, methods, and authors in the broader oeuvre as well as highlight ways in which the interfield dynamics change over time. The purpose of this examination was to provide insight into what makes an area of research transdisciplinary by identifying and examining boundary objects, or conceptual points of interest to multiple communities. This was accomplished by examining the frequencies of occurrence of concept-bearing terms measured as noun phrases in titles, and identifying the interdisciplinary spread of concepts and authors surrounding the nodes of core concepts in cognitive science. This multi- and mixed method of examination of a transdisciplinary oeuvre of science uses a strictly hierarchical classification system against which to frame the data used for analysis. This work applied methods to measure the concentration of conceptual content in a science of science study, where most authors of most prior studies examined clusters of concepts and assumed that the meaning of concepts was based on the proximity of clustering patterns.

Reasons for studying boundaries and interdisciplinarity include 1) increased interest in information diffusion, as many studies demonstrate that most innovation occurs at the boundaries of science. In this case, we must ask what constitutes a boundary. 2) Ways to translate the abstractions of knowledge from one community to another require labor and financially intensive methods. 3) Funding agencies such as the National Science Founda-

tion (NSF) have been increasingly interested in interdisciplinary research since the 1970's (Palmer and Neumann 2002). 4) Examining boundaries in information systems provides new ways to illuminate the interrelations of concepts. Understanding interdisciplinarity requires examining the conceptual, epistemic, and multifaceted dimensions of the boundaries and borders of research. Doing so requires both macroscopic and microscopic views of broader and more tightly defined community output. In this study I used an approach triangulating points of interest using a large corpus in conjunction with a bibliographic dataset found in data provided by the Web of Science, value-added NSF classification was provided by the University of Montreal.

This discussion creates the foundation for exploring the dimensions of an interdisciplinary genre of research. Unlike the oeuvre of an individual scholar or institute, the delineation between what is an area of research and what is not is in a constant state of change. In this study, a broad section of the WoS database is captured in order to accomplish the goal of examining concepts in cognitive science. For this study, data were sourced from a local copy of the Web of Science with added NSF classification. This allowed for a richer examination of overlap of concept terms in titles, and the interplay of ideas across database-defined disciplinary boundaries. As databases can only contain work done in time past, this work is a snapshot of a scientometrically defined view of cognitive science work and how it changed in the past. Efforts to predict these changes and map the state of science itself have been made, such as Leydesdorff's maps of science, but these types of efforts are not the focus of this discussion.

The identification of boundary concepts and objects and how they translate between disciplines helps make the process of interdisciplinary research easier. Despite widespread terminological standardization in the physical sciences (Strehlow 1993), many scientists will be "frustratingly vague or contradictory in defining terms" terms to describe their ideas (Szostak 2004, 222). In part, this study explores the use of terminology within the interdisciplinary area of cognitive science through the examination of high co-occurrence terms. Ideas that parallel observed phenomena in other fields may be described using similar language to bridge the reader's understanding from one concept to a new, related

concept. Conversely, terms may be polysemous—that is to say, the same term can be assigned multiple meanings.

Boundary objects and concepts are entities of interest to and acted upon by multiple social worlds. Boundary concepts are concepts that transcend disciplinary boundaries. Concepts are organized according to disciplinary boundaries that are frequently the result of long-standing traditions. A concept may inhabit multiple disciplinary spaces, and as such, is interdisciplinary as a result of its poly-disciplinary inhabitance. Such concepts transcend boundaries imposed by disciplinary-based classification systems. Classifications are created to reflect the contents of a body of knowledge, which is justified by literary warrant; in knowledge organization, this unit of analysis is known as a domain. Classification can create artificial boundaries where interdisciplinary science is done; this science transcends disciplinarily-imposed traditions.

Because the discussion by its very nature winds around itself, and each piece is highly dependent on other pieces, I outline the discussion in Chapter 1 in the following way:

1) Background of the boundary object
2) Boundaries
3) Concepts (especially as discussed in KO)
4) Classification
5) Domains, disciplines, and boundary crossing
6) Interdisciplinarity
7) Concept translation
8) Conclusion
9) Research questions and scope of the study

Questions central to this research regard interdisciplinarity and the permeability of topics, including a) the epistemological boundaries of disciplines in cognitive science, b) patterns of cross-disciplinary interactions; and c) terminology used to describe concepts.

1.1 Background

The "boundary object" was originally proposed as a conceptual mechanism for understanding interactions and information practices of researchers by Star (1988), and applied by Star and Griesemer (1989) at the Berkeley Museum of Vertebrate Zoology. The boundary object as a conceptual tool framing an entity of interest to multiple communities is used in several areas of research, and typically examines such shared entities in an institutionally-bound setting. I have taken the notion of the boundary object and use it as a preliminary basis for examining conceptual entities that are of interest to multiple disciplinarily-bounded communities of practice as the concepts and the actors working on them transcend their social and disciplinary boundaries to become encoded in language that is societally accepted, then published. These publications are then bounded by classificatory structures imposed by categorizations of science, typically done at the journal level, that impose classifications of science that are named and assumed to be disciplinary identities (such as "psychology" and "social sciences"). This discussion is grounded by providing an overview of boundaries, boundary work, and boundary objects; concepts and concept theory; classification, domains, disciplines and disciplinarity, and interdisciplinarity; concept translation; and the motivation for uniting all these ideas to address the problem of interdisciplinarity.

Star's original conception of the boundary object allows for its organic existence as an agreed-upon entity bridging multiple social worlds in the context of an institution, emerging through processes of work (Timmermans 2015). Initially agreed-upon boundary entities become something else as they are increasingly formalized between disciplines. I extend the metaphor of the boundary object to entities existing between multiple disciplinary categories in a citation database, specifically, the Web of Science. In this analysis, boundary entities examined include terminologically represented concepts and their periphery of surrounding co-occurring concepts inhabiting the conceptual space of a subset of the Web of Science. The examination of terms as they are siloed within the bibliographic database the Web of Science will allow for the examination of the theoretical and sociological underpinnings of terminological origin and the term usage by scholars.

Beginning the discussion with concepts as defined in knowledge organization creates a foundation on which to understand disciplinarily-bounded concepts. KO seeks to clarify the structure and interrelations of the products of human knowledge, and frequently does so using a domain-analytic paradigm that delineates what is of a community from what is not of a community by defining the intension and extension of a domain. Communicating concepts as they are understood by a community requires the creation of systems to maintain and convey the collective understanding of ideas and their relations, as "like things" ought to be grouped together (Hjørland 2003, 87); however, the consideration of what is similar to some other thing "is not a trivial question" (87). These relationships must be codified in a way that can be understood and reused by others. The interrelatedness of concepts can be clarified through their abstraction and representation in a classification system.

Suppe (1989) claimed that classification belongs naturally to the use of language, and thus to communication. He referred to the use of predicative phrases to logically group actions or properties in like-kinds as "conceptual classification," which is a broad sense of classification (292). This distinction of a broad type of classification is intrinsically phenomenological, as it refers to an individual's descriptions of their experiences and surroundings. Classification is "the process of determining where an information resource fits into a given hierarchy and of then assigning the notation associated with the appropriate level of the hierarchy to the information resource and its surrogate" (Taylor and Joudrey 2009, 448). The formal representation of the understanding of a community of the qualities and characteristics of its own knowledge bound into a cohesive structure separating concepts into non-overlapping classes Jacob (2004). The unit of measure for the classification of communities of practice is typically discussed as a "discipline."

Both communities of practice and areas of research are described as disciplines; thus their discussion logically follows. From disciplines, I outline types of disciplinary linkages, or measurable and unmeasurable social exchanges between formalized communities of practice; here I focus on the social nature of disciplines because disciplines are social. As the nature of problems in question change, the habits of actors in a discipline. This

results in acts of borrowing and collaboration outside of an actor's home discipline. Interdisciplinarity requires socialization across disciplinary, institutional, geographical, and otherwise socially imposed boundaries. The discussion of these social exchanges sets up the discussion for interdisciplinarity, followed by interdisciplinary information needs and seeking. Although interdisciplinary information needs are not the crux of this discussion, understanding these needs provides insight into the shortcomings in classification and information retrieval practices. Disciplinarity, interdisciplinarity, and disciplinary communication build a basis for the discussion of boundaries, boundary crossing, and boundary objects. The interrelatedness of these ideas is important because they build on each other in meaning, and their interconnectedness works to create a better understanding as they continue to be examined.

1.2 Boundaries

A "boundary" is any delineation between two or more entities, and can be thought of as a division between what an entity is and what it is not. Boundaries themselves serve to create natural or artificial divides in multiple types of landscapes. Some of these divisions are natural, such as geography or researcher experiences, while some are artificial, such as the institutionally-imposed structure of the academic unit or classification. Boundaries between scientific disciplines must be identified in order for "boundary work" to progress (Szostak 2004, 173). From a sociological perspective, Bowker and Star (1999) discuss individuals who engage in this work as boundary spanners who occupy the "borderlands" between communities. Individuals who occupy the borderlands of research frequently must work to make their views heard, as the synthesis of disciplinary views may be less seriously regarded in scholarly communities. These researchers may bridge two or more research communities, but can be seen to belong to none. Clear and precise definition of disciplinary boundaries makes the linking of multiple scientific specialties easier Szostak (2004).

Szostak (2004) outlined historical interest in the unification of science in his discussion of classification of scientific practice. Philosophers of science, discussed in the introduc-

tion, emphasize the need for individuals and communities to develop their understanding of "what" into language and shared discussion. Disputed boundaries and objects could be handled through representation of approaches and methodologies used to examine shared phenomena of interest, ensuring that disciplinary perspectives are represented in such contentious cases (Hjørland and Pedersen 2005, Szostak 2008).

Researchers experience boundaries present in literature through requirements for publication of our work, database limitations in the form of imposed classificatory structures in representing and storing the content of the work we have created, and others. Researchers must engage in boundary work in all arenas of work as they mediate navigation of complex subjects. Boundary work as discussed in multiple fields works through different levels of individuals, situations, and institutions. These boundaries can be navigated and explored through examining boundary work, which can include boundaries in polycontextual situations (Engeström et al. 1995), to remedy individual identities against what they know of themselves and how they communicate with others (Postholm 2015), sensemaking between research contexts (Weick 1988), remedying the interdependency of conflicting knowledge (Carlile 2004, Howard-Grenville and Carlile 2006), framing compromise in organizations (Jeantet and Vinck 1995), and to stabilize facts across social worlds (Fujimura 1992).

Boundary objects 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 (Bowker and Star 1999). Each boundary object has different meanings to members of distinct communities who act upon the object (Star 2010). Examples of formalized boundary objects include ontologies, and metadata crosswalks; however, boundary objects can potentially enhance cooperation, coordination, and knowledge management across different disciplines involved in scientific research. Griesemer and Star define the notion of a boundary object as follows (Star 2010):

- The object (remember, to read this as a set of work arrangement that are at once material and processual) resides between social worlds (or communities of practice) where it is ill structured.

- When necessary, the object is worked on by local groups who maintain its vaguer identity as a common object, while making it more specific, more tailored to local use within a social world, and therefore useful for work that is NOT interdisciplinary.
- Groups that are cooperating without consensus tack back-and-forth between both forms of the object. (604-605)

Objects residing between multiple social worlds maintain meaning for multiple communities. This meaning of an object can be expanded intentionally, so the object has more and more obvious meaning to members of multiple communities, or it can be tailored to specific needs. The tailoring of boundary objects to specific community needs strips external perspectives from the meaning imbued in the boundary object, thereby limiting its interpretive value already existing within the community.

Moving toward the abstract, boundary objects can themselves be concepts. "Boundary concepts" are concepts that span the social, cultural, and theoretical boundaries of two or more disciplines. As such, framing boundary objects as boundary concepts provides a more flexible means of analysis, as it allows for the negotiation and synthesis of emerging shared areas of interest into an active area of exchange, following a social constructivist arc (Allen 2009).

1.3 Concepts

For the purposes of this work, a "concept" is defined as a linguistically expressed unit of understanding that can inhabit shared intellectual space, as described in Hjørland's (2009) "post-Kuhnian" view of scientific paradigms. The notion of "post-Kuhnian" paradigms is a more realistic view of science in that scientific paradigms can inhabit a shared intellectual space, where Kuhn's (1962) theory of paradigmatic shifts accounts for a view of science that can be described as Darwinian in nature. Concepts can be used to break down and quantify the worldview of individual researchers, as is implied by the phenomenological overtones of Suppe's (1989) conceptual classification. Phenomenology provides a basis for understanding how an individual interprets the world, and what influences went into shaping their worldview (Husserl 1950). The individual's ability to relate to ideas encountered, and then to incorporate them into personal experience is described

as the process of "empathy" (92). The influence of a researcher's education, experience, and work environment will influence how each individual conceptualizes, synthesizes, expresses, and disseminates the created knowledge in formalized discourse as it has been distilled and reinterpreted according to an understanding developed by their individual experiences. As noted by Hjørland (2009), "knowledge organization systems … should be considered to organize collections of concepts."

Concept theory, as discussed in knowledge organization, is our understanding of how people conceptualize and learn new things (Dahlberg 2006, 2009). Theories play a central role in the development of the human understanding of ideas (Hjørland 2009), furthermore, they are vital to the human ability to differentiate, classify, and relate competing concepts. Hjørland (2009) argues that theories of concepts can be shaped by the major pillars of epistemology through which they are examined (empiricism, rationalism, historicism, and pragmatism) as they provide a lens through which to examine understanding of the world and the types of research that are conducted in order to better understand it (Smiraglia 2008). From rationalist and empiricist perspectives, concepts represented in information systems are considered to be objective units of knowledge; from a historic or social constructivist standpoint, concepts in information systems are representative of cultural, domain, and individual/phenomenological processes that developed from their examination and use (Albrechtsen and Jacob 1997).

These four pillars provide a means for the classification and understanding of concepts, and thus form their own concept theory (Hjørland 2009). Szostak (2010) stresses that concepts can be shared across communities, and representing complex concepts as simpler concepts does not detract from the importance of a theory or phenomenon to a particular discipline; instead, it makes the component concepts more translatable between communities with shared interests. The analysis that is conducted in this dissertation is empirical in that it is evidence-based, rationalist in that it examines knowledge in both a priori and a posteriori manners through the examination of classification (a priori) and natural language (a posteriori), and historical in that it uses a corpus of work on cognitive science that was been done in the past.

Communication in general can be understood through revisiting the roots of semiology and symbology. Peirce (1992) introduced a threefold model of semiology including a representamen, object, and interpretant that are used both by the sender and receiver of a message; similarly, Saussure's (1959) semiotics used a dichotomous representation of the sign itself as its meaning passes from signifier to signified. Though knowledge may have a specific meaning intended when it is first encoded to be shared, its meaning may be altered or reinterpreted by the passage of time, reception by other communities, or an individual's process of reaching understanding. Concept terms may have multiple, or polysemous, meanings that can cloud user interpretation, further complicating matters of sharing knowledge. Communication is rooted in several disciplines. For the purpose of this work, I draw on both the theoretical and empirical. The theoretical groundwork laid in semiology and symbology are closely tied to the foundations of knowledge organization. Each sign bears an intended meaning steeped in the traditions of its origination. When it is shared, the meaning can become mutable based on the context. Once meaning has been altered, multiple (or polysemous) meanings for the same symbol, or term, can occur in multiple contexts. This multiplicity of sign interpretations can lead to confusion, especially for researchers whose background may differ from the context of the embedded sign.

Language is our primary means of sharing knowledge through time and space (Dahlberg 2006). Furthermore, the words used to carry meaning are put in a specific grammatical order in order to make sense (Collins 1998, 858-859). Scientific work as it is published is encoded knowledge in a way that makes it sharable and reusable across time, space, and multiple communities. Upon entrance into a new community of practice or field of research, an initiate is indoctrinated in the argot of their new domain through "learning-as-membership" (Bowker and Star 1999, 295). Confusion and uncertainty can occur in information seeking when newly acquired knowledge is described by repurposing concept terms from other disciplines (Ridenour 2015).

1.4 Classification: From concepts to Shared Understandings

Classification, by definition, provides a means of representing the relationships between ideas as they are shared by and discussed within a community (Bowker and Star 1999). Classification should expedite access to information contained in retrieval systems; it should be noted that classification as it has been found in library systems serves to represent a disciplinary identity (Bates 1996). These disciplinary identities do not necessarily neatly align to the research conducted in that area, as the problem may be shared across disciplinary boundaries. Storing knowledge in a way that expedites access for those indoctrinated into the traditions of one discipline creates barriers to access for those new to the research within a domain. This results in "silos" often discussed by scholars of information science. The boundaries imposed by classification in the information system pose barriers created by the delineation of categories that do not necessarily represent the manner in which the research was conducted. This is especially true of inter- and transdisciplinary science, where scholars from multiple backgrounds work together to solve shared problems. Classification can be used to communicate and represent shared understandings of what is known, but it needs to provide information in a way that is conducive to cross-disciplinary boundary work. For this to happen, library services must be able to accommodate classification for multiple disciplines (Palmer 1996).

Dervin (2003) discussed the current classification practice of what (Hjørland 2002, 22) termed "psychologizing epistemology," emphasizing the need to "epistemologize psychology." Though taken from its original context, the psychologizing of disciplinary epistemology results in classifications heavily imbued with disciplinary identities, which complicates classifying knowledge in ways that it can be found by information seekers of diverse epistemological backgrounds.

1.5 Domains, disciplines, and boundary crossing

In knowledge organization, Smiraglia (2014) succinctly defines a domain as "a group with an ontological base that reveals an underlying teleology, a set of common hypotheses, epistemological consensus on methodological approaches, and social semantics" (114). The

unit of the domain provides a functional unit of analysis for analyzing the aboutness of a community of practice. It is important to note that the scope of a domain is determined by the researchers conducting an analysis, and may vary. The identification of boundaries for the analysis of domains were identified by Tennis (2003); these include their intension, or granularity, and extension, or coverage in time and space. KO uses the unit of the domain as the basis for investigation in order to construct knowledge organization systems that reflect the understanding of cohesive groups of some sort. Domains themselves are not necessarily communities; however, any self-identified community may be analyzed as a domain. Other community units are discussed throughout this work; next is a discussion of disciplines and what distinguishes a discipline from a domain, and from an interdisciplinary area.

Academic domains are often referred to as "disciplines" (Smiraglia and Lee 2012, 15), and these domains consist of cultures involving distinct vocabularies related to describing shared theories. Disciplinarity is a social phenomenon often driven by a sense of tradition (Sugimoto and Weingart 2015). I define disciplines as self-contained research entities with established traditions made up of scholars who engage in self-similar implementation of shared conceptual structures in information seeking practices, encoding of concepts in literature, and submission of produced knowledge to journals. The definition alludes to the often isolating nature of the bounds of disciplinary entities. Sugimoto and Weingart (2015) highlight that for an area of research to be a discipline, the area must possess a state of self-identity that can be either reinforced or broken down over time. For an area of research to be considered a discipline, that area must have sufficient investment in the forms of recognized scholars, institutional recognition, funding, and other social capital factors.

Disciplines also involve the development of tradition, which is a process that takes a great deal of time and goes through several stages of development. These stages of disciplinary formation vary in description, but in general, more matured disciplines are better structured. As disciplines evolve and new knowledge is produced, new disciplines are then formed to accommodate the evolution of scientific understanding of the universe

Zhang and Jacob (2013). Inchoate disciplines may lack sufficiently developed terminology with which to describe their ideas in ways that can be utilized for controlled vocabularies Hjørland (1997). What (Hjørland 1997, 138) describes as "ill-defined disciplines" parallel (Kuhn 1962, 48) immature sciences, or "pre-paradigms," in which lack of agreement upon theories, methodologies, and procedures results in chaotic communication. Galison (1995) viewed science and disciplinarity as a series of island nations, which is similar to Kuhn's view of scientists who work and communicate inside of their chosen paradigms. Classification for the knowledge contained in the bodies of work published by these groups may be simple, and thus described as "naïve classification" by Beghtol (2003).

As this work uses a text-based sampling for the inclusion of journals as the initial unit for the core of the domain, the analysis conducted is a domain analysis and not an analysis of a discipline. No presumption of tradition was made for the inclusion of particular journals, in part as this was an exploratory study.

Invisible colleges are groups of researchers whose unofficial affiliations are transitory in nature, but shape each researcher's output (Price 1963). Members of these groups influence one another based on exchanges in their shared collaboration spaces. Crane (1969) argued that invisible colleges are an example of a social group formed in response to a need for specialty knowledge. Kuhn's (1962) scientific paradigms pre-dates Crane's, but clusters scientists into dedicated groups working to solve a prescribed set of problems.

Growth in scientific disciplines can be examined on multiple levels and from multiple perspectives; Kuhn (1962) held a view of paradigmatic shifts in disciplines that evolved as a response to changes in the collective understanding of science, while de Solla Price suggested that scientific specialties would undergo speciation when a threshold of more than a hundred active members of a discipline was breached (Price 1963, Leydesdorff 2006). Kuhn's shifts in paradigms have been subject to question in recent years, as a more sociological focus to understanding the dynamics of science and the migration of individuals and ideas.

Sugimoto and Weingart (2015) outlined various aspects of disciplinarity, including conceptualizations of disciplines (the cognitive, social, communicative, "separatedness,"

tradition, institutional, and combinations of the above discussion), narratives demonstrating the maturity of disciplines (the great man, the establishment of societies and conferences surrounding an area of interest, governmental funding and recognition of the discipline, a social need for the area of interest, institutional recognition of the area, publications documenting knowledge within the area, and the relationships to other disciplines). Broadly, research fields are areas of inquiry that have multiple researchers who are working together to solve a problem (Klein 2000).

Disciplinary boundaries must be established, defined, and quantified in order to meaningfully compare concepts between and across disciplinary boundaries (Szostak 2004). Boundaries are social, physical, geographical, or traditional distinctions between what is one discipline and what is not. Once boundaries have been articulated, examining cross-disciplinary patterns of communication around problems as they are of interest to multiple areas of research is possible.

Cross-disciplinary exchanges are frequently referred to using terminology that is often treated as interchangeable; these terms, cross-disciplinarity, multidisciplinarity, interdisciplinarity, and transdisciplinarity, have distinct meanings that describe the degree of interdisciplinary integration between and across disciplines. Klein (2010) defined a spectrum of degrees of interdisciplinary integration ranging from multidisciplinary to interdisciplinary to transdisciplinary. In her proposed system, based on the underlying notion of disciplinarity and disciplinary identity, "multidisciplinarity" is the least integrated form of cross-disciplinary sharing, as disciplinary specialties retain individual identities and individual disciplines' scholars work to compliment and build upon the knowledge they contain through juxtaposition and coordination of knowledge within other disciplines. Thus, interdisciplinarity is the result of integrated coordination and directed collaboration efforts; and while transdisciplinarity is the highest level of cross-disciplinary exchange. Transdisciplinarity becomes "simultaneously an attitude and a form of action" (Klein 2010, 521) in that a problem or method transcends the scope of one discipline and becomes a "common system of axioms that transcends the narrow scope of disciplinary world views through overarching synthesis" (Klein 2010, 24).

Frameworks for understanding cross-disciplinary interactions have been created by various scholars of interdisciplinarity. Bechtel (1986) identified five patterns of cross-disciplinary linkages, where Vinck (2000) described for models of interdisciplinary interactions (complimentary, circulation, fusion, and confrontation). Bechtel's five patterns of cross-disciplinary connections consist of (1986, 46-47):

- Developing conceptual links between disciplines to adopt and modify perspectives from one to the otehr without overwhelming the adopting discipline with theoretical structures
- Recognizing a new level of organization (such as a new field or theory) in order to solve unsolved problems in existing fields
- Adoption of techniques and methods from one field to another to help build on theories in the adopting discipline
- Modifying and extending theoretical frameworks from one domain to another
- Development of new theoretical frameworks to integrate and synthesize research from separate domains

Vinck's four models of interdisciplinarity serve as starting points from which other hybridizations of interdisciplinary actions may occur (2000, translated):

- "Complimentarity" is the result of actors prioritizing the joint examination of a problem. It is typical for one of the involved disciplines to be more involved than others, resulting in asymmetrical participation
- "Circulation" is the result of actors in a discipline borrowing concepts, methods, questions, or problems from one or more other disciplines
- "Fusion" is the regrouping of researchers to address a shared problem while abandoning previous disciplinary identities
- "Confrontation" is the result of negotiating the crossing and redefining of boundaries as is necessary to honor strongly held disciplinary views

Both the frameworks created by Bechtel (1986) and Vinck (2000) provide bases for discussing the needs of interdisciplinary scholars based on the type of cross-disciplinary connections that may be made. Conceptual links must be identified across field boundaries; fields, theories, theoretical frameworks, models, techniques, and methodologies must be explicitly identified and indexed in knowledge organization systems.

Communication between communities implies discourse (Greisdorf 2000), and as such, examining cross-disciplinary communication can reveal potential relations that could be incorporated for a more thorough interdisciplinary framework for document classification. Communication breakdowns occur between specialties and not between individual researchers (Wilson 1993). Having established a framework for understanding disciplinary interactions, I now discuss what constitutes interfield exchanges between disciplines.

Scientists from multiple disciplines who work together to solve problems and to create knowledge are frequently referred to as interdisciplinary teams. Philosophically, the testing of hypotheses is independent of socially imposed disciplinary constraints, which has been recognized by philosophers of science and interdisciplinary scholars for many years. Darden and Maull (1977) introduced the idea that a theory could work to bridge fields, referring to such theories as "interfield" and outlined the necessary conceptual (as opposed to sociological) functions of problems and facts gathered to solve them. In this case, the spanning theory becomes boundary object, as it becomes a theory that is held within conceptual boundaries of multiple fields of science.

1.6 Interdisciplinarity

Defining interdisciplinarity is a difficult task, as so many views contribute as many (or more) definitions of interdisciplinarity and interdisciplinary research. Porter and Rafols (2009) give interdisciplinarity the mission of "advanc[ing] fundamental understanding or to solve problems whose solutions are beyond the scope of a single field of research practice" (2009, 720). From a bibliometrics perspective, Garfield, Malin, and Small describe interdisciplinary research as "linkages between specialties of diverse subject patterns" (1978, 189). Each field's perspective and understanding of what interdisciplinarity is

contributes unique insight and methodologies nuanced by the epistemological underpinnings of each group's beliefs.

What constitutes interdisciplinarity has been of interest to researchers in many fields for a long time. Wagner et al. (2011) conducted an extensive literature review of participant views of interdisciplinarity and conditions that must be fulfilled in order to consider research interdisciplinary. Functionally, interdisciplinarity can be seen as a type of hybridization of disciplines. This hybridization can lead to the formation of a new discipline Milojević (2009), but the steps made in cooperation and integration leading up to the creation of a new discipline can be analyzed in various stages. Klein (2000) outlined a vocabulary for discussing interdisciplinary work, in which she discussed the linkages between disciplines as cross-disciplinary linkages (discussed above). In order for research to be considered interdisciplinary, it must contain citations to multiple Essential Science Indicator (ESI) categories (Moed 2015). Considering these definitions, interdisciplinary research is research that spans, is influenced by, and contains the work of diverse scholars from multiple disciplinary backgrounds.

Teams of scholars and scientists who come together to solve problems must seek information outside of their own comfort zone, be it through social communication at conferences, databases containing unfamiliar terminology, or literature describing problems of interest in ways novel to the individual. Inherent to interdisciplinary work is the multiplicity of perspectives, philosophies, and approaches contributing to the creation of new knowledge. Research done to build on inter- and transdisciplinary ideas requires searching in order to build connections between what is known and what is needed in order to articulate the relations between fields. The act of searching relies on serendipity, or the moment of making hidden connections between what a searcher knows and what they need to find (Foster and Ford 2003).

Research areas of broad interdisciplinarity, such as information science, cognitive science, and informatics, have grown from multiple parent disciplines in which research is still actively conducted. Contributions to these interdisciplinary fields all work to address similar issues but contribute different disciplinary perspectives to the problem being ad-

dressed. For example, disciplines that contribute to cognitive science include psychology, biology, philosophy, neuroscience, linguistics, and anthropology, among others. Contributions to this field all address cognition in some way, but from different theoretical backgrounds and disciplinary perspectives. Information needs of scholars in this area may include the ability to trace contributions to a field from a contributing discipline.

Classification at the information-system level creates boundaries not present in the social world, and must be structured to facilitate the inclusion of inter- and multi-disciplinary perspectives in a way that is conducive to cross-disciplinary boundary work. For this to happen, library services must be able to accommodate classification for multiple disciplines (Palmer 1996). Dervin (2003) suggested that working toward a methodology for interdisciplinarity would ultimately lead to better ways to collaborate, instead of creating lists of unrelated data, phenomena, classifications, techniques, and methodologies. The need to identify and pursue "causal links" between related phenomena is obfuscated by the tendency of scientists to use "frustratingly vague or contradictory" definitions for terms (Szostak 2004, 222). Szostak's recommendations for moving to a phenomenon-based classification system provide one potential solution for solving the problem of interdisciplinary classification and information retrieval; however, his proposal has been considered radical and has met with resistance. Pursuing the classification of causal links would, however, address the call to "cease the production of unrelated facts" (Hjørland 1996, 52).

1.7 Concept Translation

In 1972, Popper discussed a second world, the mind of the individual, and a third world, which contains abstract products of thought such as scientific theories. In his model, the process of learning and an individual's inquiry in the third world can change what is contained in the second world, or personal mind. Popper (1972) went on to describe two thought experiments demonstrating the objectivity of knowledge. In the first, the subjective learning and knowledge contained in the third world, are destroyed, but libraries and the individual's ability to learn, creating the second world are maintained. In the second,

the libraries are destroyed, in addition to the tools of civilization, rendering our ability to learn from them irrelevant. This second thought experiment drives home that our ability to learn from the third world is dependent on the maintenance of outside repositories of knowledge, such as libraries, and the very fact that these repositories exist independently of the individuals who created individual pieces of knowledge. Subjectively, the third world can gain meaning based on our understanding of the communities who created it; objectively, the third world exists apart from the traditions of the communities who created it. With regard to disciplinarity, the strictures imposed by indexing practices with regard to the communities who created the knowledge inhibit its ability to be accessed by a diverse audience.

Given the aforementioned discussion outlining concepts, concept theory, disciplinarity, and disciplinary boundaries, it now makes sense to discuss concept translation. Shared conceptualizations between domains may be represented by different terminology reflecting each community's epistemological stances. This terminology is selected to bridge concepts and problems, but is inherently unstable due to changes in understanding (Courtial and Law 1989). Such shared conceptualizations as they bear meaning to multiple disciplines are referred to by various names; in general, I refer to them as boundary concepts, or boundary objects (Bowker and Star 1999).

1.8 Conclusion

In summary, interdisciplinarity is a complex phenomenon. Current understandings of interdisciplinarity do not result in the creation of classification that is useful to describe the work done in such an area, or in a way that is useful for interdisciplinary scholars. Boundary objects and concepts inhabit, or are acted upon, multiple communities and their members in such a way that tacit knowledge regarding the surrounding aboutness of terminology may be lost in the context of a given classificatory system. This work illuminates boundary concepts contained in titles against the classificatory scaffolding of the National Science Foundation on a large bibliographic dataset using methodology tailored to identify boundary crossing concepts. This novel set of methods can be ap-

plied to other interdisciplinary bodies of work, providing guidelines for the detection of boundary-inhabiting entities in complex information.

1.8.1 Research Questions

In order to examine interdisciplinarity in a domain analysis from the perspectives of what (topicality), who (authorship), and when (time), methodology was crafted based on existing methods to answer the following research questions:

1) What epistemological boundaries, manifested as classificatory divisions, exist between NSF-identified disciplines that comprise cognitive science? What is the representation of work published in these? How does this change over time?

   A) How do terminological boundary objects fit against the NSF's classificatory framework? What concepts are the most interdisciplinary in the core of the domain?

   B) What boundary crossing concepts are of the highest frequency? What boundary concepts at the core-level are 100% interdisciplinary, and what are their topical periphera? How do boundary-crossing concepts in the core change over time?

2) What are the underlying sociology (through acts of co-authorship)? What acts of author boundary crossing exist surrounding the most interdisciplinary concepts?

1.8.2 Outline of the Dissertation

This work is outlined in the following way:

- Chapter 2: an overview of research on boundary objects in the areas of sociology' social network analysis; medicine; education and psychology; management, organizational science, and information systems; and information science. Chapter 2 discusses the relationship between boundary objects and interdisciplinarity, as the

two are closely linked by the nature of the boundary object serving as a point of translation between two or more communities of practice. Following the discussion on boundary objects and interdisciplinarity, different measures of interdisciplinarity are outlined.

- Chapter 3: methodology used for analyzing text, authorship, and bibliometric phenomenon are discussed. Most methods of automatic text processing applied to corpora result in either a model of statistical fit or a projected number of ideal classes based on calculations made by an algorithm; as the questions in this work delve into topicality itself, none were deemed appropriate for application here. However, they must be mentioned as they are the basis for much work done in textual analysis. Informetric and bibliometric methods for examining citation-based relationships and their resulting analyses, as well as hybridization of methods are also discussed.

- Chapter 4: an outline of methods used to converge on highly interdisciplinary boundary-inhabiting, or boundary spanning, concepts in the core domain of cognitive science. The study design is illustrated in a methodological pipeline designed to filter data to subsets that were appropriately interlinked in order to examine concepts, how concepts changed over time, the authors who published work on those concepts, and measures of interdisciplinarity adopted for the results. As transparency was a primary motivator for this work, simple metrics were adopted to allow for an in-depth examination of what was being published about and when it was being published by whom.

- Chapter 5: a discussion of the results of the study, including identifying the taxonomic overlap, which is presumed to be an indication of epistemological boundaries because of the nature of classification and disciplines. Boundary crossing, or boundary spanning, concepts are outlined and presented in temporal visualizations for both the entire transdisciplinary domain, as well as the core of the domain. The topical periphera of the most interdisciplinary concepts ("children's"

and “case study”) are visualized using a co-occurrence network, as the linkages between co-occurring terms result in clusters of related concepts that themselves were not necesarily boundary-crossing in nature. However, the periphera provides context for how each boundary spanning term is studied (the methods), and research problems related to each concept. Evidence of established underlying sociology, as demonstrated through the act of publication, is discussed for the most interdisciplinary core topics (in this case, “children’s,” and “case study”).

- Chapter 6: a discussion surrounding the assumptions made in this work, the nature of the data, and the nature of text-based data and pre-packaged tools. Also discussed are different ways of examining concepts in natural language, and the choice to use whole noun phrases instead of stemmed or lemmatized word forms. Potential applications of this type of analysis are also included in this chapter.

- Chapter 7: a discussion of future work involving the notion of the boundary object and text processing. Revisited are Marchese and Smiraglia’s (2013) idea of the pivot point, and how this could be applied to text-based boundary phenomena. A discussion of collocating concepts is included, as collocations may, in fact, be more useful devices for determining differing viewpoints and their representation in published literature.

Chapter 2

Literature Review

2 Literature Review

Boundary objects are traditionally studied in institutional settings using qualitative methodology to gather data regarding the objects and the multiplicity of views surrounding their polycontextual existence. Bateson (1972) famously stated that "information ... is a difference that makes a difference" (99). In the context of boundary objects, understanding the differences between distinct community and individual views involved in the creation and maintenance of a boundary object itself allows for analyzing possible interpretations of said boundary object. Huvila et al. (2016) examined boundary objects as they have been discussed in the field of information science.

In this review of literature, research on boundary objects is discussed from interdisciplinary perspectives. From boundary objects themselves, an outline of how classification and indexing are interlinked, and distinct processes of indexing and thesaurus construction are discussed. These are relevant to this discussion because term selection for describing works of interest to multiple audiences is problematic.

2.1 Research on Boundary Objects

A boundary object is an abstract or concrete entity that is shared by multiple communities of practice. Though the concept of the boundary object originates in sociology, the notion of a shared idea interacted with by multiple entities has been widely investigated by many names in multiple fields of study. These objects require shared social semantics over items

or ideas whose domain-independent meanings may not neatly align across disciplinary boundaries. Boundaries can divide knowledge itself (Carlile 2002). In organizations, these divisions can be seen as localized, embedded, and invested.

In order to understand boundary objects, we must examine factors that contribute to the creation of and maintenance of boundaries. Types of boundaries discussed include syntactic, semantic, pragmatic, social, and symbolic. Carlile (2002) outlined the functions of syntactic and semantic boundaries as found in the literature, and proposed pragmatic as a third type based off of ethnographic research surrounding four primary communities with distinct functions in a product-developing organization. According to Carlile (2002), Shannon and Weaver's mathematical theory of communication is at the syntactic level of boundaries, in that a unit of communication such as a code is subject to the interpretation of the sender and the receiver (Shannon 1948). Semantic approaches to boundaries recognize that individual and collective interpretations of common symbols result in unique differentiation of understanding that hinder collaboration. Carlile's proposed pragmatic approach examined knowledge across the four distinct communities of practice examined knowledge as it was situated locally "in character," meaning that similar knowledge in one function's arena can be applied to similar problems in another arena.

Factors that shape boundaries between disciplinary and sociological factions of organizations vary. Boundaries such as race, gender, class, and territorial inequality are referred to as "symbolic boundaries", as their conceptual distinctions are made on the basis of categorization of traits done by social actors (Lamont and Molnár 2002). Symbolic boundaries that become widely accepted and agreed upon can begin to constrain social behaviors, creating socially acceptable and acted-upon divisions that become apparent in examination of the system, and result in "social boundaries" (169). Lamont and Molnár advocate for the examination of both symbolic and social boundaries as equally real divisions; symbolic boundaries are intersubjective in nature but "manifest as groupings of individuals. At the causal level, symbolic boundaries can be thought of as a necessary but insufficient condition for the existence of social boundaries" (Lamont 1992, 192).

2.2 Disciplines that study and discuss boundary objects

Boundary objects are themselves discussed in multiple disciplines under many different names. For the purposes of this discussion I refer to research areas and disciplines by high- level field names, as opposed to highly specific disciplines and specialties. This high-level delineation of the discussion permits a greater synthesis despite potential segregation of communities. Each field views and studies boundary objects from its own epistemological perspectives, nuancing the levels at which it examines boundary objects as identified in their respective areas of interest. Fields included in this discussion include sociology, social network analysis, medicine, education and psychology, management and organization science, information systems, and information science.

2.2.1 Sociology

In sociology, boundary objects are frequently discussed in the context of institutional ecology. Fundamental to the sociological understanding of boundary objects is their shared nature, as such objects occupy a shared social space in which members of multiple communities of practice interact. Star and Griesemer (1989) coined the term "boundary object" (387) in 1989 when examining "translations" (389) between professionals and amateur naturalists and patrons in the Museum of Vertebrate Zoology at Berkeley. Such objects inhabit a conceptual, cognitive, and technical space and are flexible enough to be adapted to multiple points of view. Boundary objects reflect the "fundamental tension of science" (Star and Griesemer 1989, 392), as they are created in and occupy multiple social worlds. The objects have different meanings in the social worlds that are "sufficiently structured" to be recognized by all worlds (Trompette and Vinck 2009, e). Latour's (1987) work on network analysis influenced Star and Griesemer's work in which they examined both the "flow of objects and concepts through the network of participating allies and social worlds," effectively shifting the focus from a network model to organizational/localized collective activity 1989, 389.

Fujimura (1992) advocated for use of the term "standardized packages" (203) as a concept to mediate activity across and between social worlds and to provide "fact sta-

bilization" (204), which she later described as "fact (and skill) stabilization" (Fujimura 1996, 152). Standard packages serve as a sort of meta-boundary object, combining both multiple boundary objects and sets of standardized methods. Where Star and Griesemer's boundary objects provide a theoretical or practical construct for collective understanding and action, standard packages offer consistent shared theories and methodologies for multiple communities.

2.2.2 Social Network Analysis

Social network analysis can illuminate the roles of actors (or nodes) by examining their positions and connections to other nodes in the network Easley and Kleinberg (2010). The nodes in a network with the greatest potential for knowledge exchange are those positioned between homogenous clusters, creating zones of heterogeneity that bear the potential for the exchange of new knowledge (Reagans and McEvily 2003). Conversely, structural holes in a network are places in a network where clusters of relationships break off, and are bridged by individuals who mediate between two or more clusters (Reagans and McEvily 2003). Mediating individuals serve as boundary spanners and engage in boundary crossing. Nodes in boundary-spanning positions, known as weak ties, are positioned between multiple groups. This positioning allows intermediary actors to have access to greater knowledge and the means to transfer it to other cliques in the network.

2.2.3 Medicine

In medicine, the study of boundary objects is closely tied to sociology. Objects themselves range from physical items such as paper charts and electronic health records to concepts and terminology used in both formal and informal communication. Since 1951, interdisciplinary care teams were tasked with cooperation despite differences in academic rank, socioeconomic status, and popular publicity regarding revolutionary therapeutic techniques (Caudill and Roberts 1951). Allen (2009) described the appeal of boundary concepts, especially in medicine (355):

> A boundary concept is a loose concept, which has a strong cohesive power. It is precisely because of their vagueness that they facilitate communication and cooperation between members of distinct groups without obliging members to give up the advantages of their respective social identities.

The ability to communicate meaning without abandoning an individual's personal and occupational identities allows for cooperation without a sense of betraying oneself. In many fields, the individual's occupational identity comes with both perceived expertise and rules governing acceptable behavior and decisions that they are allowed to make. In this manner, the tension of science discussed by Star and Griesemer (1989) extends to the occupational identities of individuals.

Instances such as the coordination of care teams are a good example of the multiplicity of educational and professional backgrounds found in medicine, as collaboration required from professionals with distinct training is diverse. Doctors, nurses, surgeons, physicians, patients, managers, and insurance specialists must all communicate effectively to treat patients, making the examination of boundary objects ideal for understanding how these teams communicate in their respective systems. The separation of the self from the occupation, as well as the ability to navigate multiple social worlds, is closely tied to the framing of boundary objects in education and psychology.

2.2.4 Education and Psychology

Education and psychology examine boundary objects as they pertain to the individual student's cognitive processes; navigation of multiple social worlds; developmental processes and their relation to culture; multiplicity of contexts, or polycontextuality, in learning and instruction. They also look into remedying the spanning of boundaries with the increasing globalization of education.

Boundary crossing can occur for students when they navigate the sociocultural norms as imposed on the student as they progress through their education, and even between the values held by teachers of different subjects (Akkerman and Bakker 2011). Akkerman and Bakker (2011) discuss learning from a sociocultural stance surrounding identity

development. This in itself, though not explicitly stated, reflects an underlying phenomenological way of thinking, as "a key question is the distinction between what is a part of me versus what is not (yet) part of me" (132).

Examining outcomes and goal-directed actions, Postholm (2015) discussed "mediating artefacts" in education using a Cultural-Historical Activity Theory framework to outline developmental processes (CHAT). Mediating artefacts associate local activities to culturally shared historical and cultural connections (Wertsch 1993). CHAT seeks to provide a framework for the developmental process of the individual's as founded by and integrated through social, educational, and historical activities in which the individual participates. Yeo and Tan (2014) discuss artifacts that are capable of mediating boundary crossing by bridging different activities; boundaries include time, space, sociocultural situations, and independent cognitive processes of students.

Engeström et al. (1995) discuss polycontextuality at the level of tasks, work actions, and activity systems (such as a community of practice or an institution) in the area of learning and instruction. Polycontextual situations occur when experts move between multiple situations in multiple context that "demand and afford different, complimentary but also conflicting cognitive tools, rules, and patterns of social interaction" (Engeström et al. 1995, 320). Engeström, Engeström, and Kärkkäinen (1995) describe "expertise" as working, or practical intelligence. A student's intelligence is partially a measure of how well they translate concepts and skills from one area to another. In this way, boundary crossing can be viewed as an act of the individual remedying internal conflicts between polycontextual situations.

As boundaries can cause potential difficulties in translation of views between cultures, they can be viewed negatively despite the process of crossing them providing individuals the opportunity to refresh their views and assumptions regarding their own practices (Tsui and Law 2007). Globalization and its impact on higher education presents unique challenges for teacher accreditation, as globalization itself has raised questions regarding how we regard knowledge (Tsui and Law 2007). Boundaries in international education are multidimensional, spanning international and local economics, policy, and practice. A need

for globally recognized accreditation and licensure for teaching has arisen partially from the current focus of economic outcomes of the "investment" of learning. School-university partnerships involve boundary crossing mediated by a boundary object of lesson study, which itself is referred to by different terms in different languages and communities of practice. Students may eventually find themselves in positions where they must translate skills learned in one area of study to another field in order to present themselves as viable candidates for positions of employment, requiring them to understand how to translate their tacit knowledge and skills from one community to another.

2.2.5 Management, Organizational Science, and Information Systems

Literature regarding boundaries and boundary objects in management and organizational science relate to the use of space, individual roles in collective activities, the evolution of the understanding of meaning, and events. Though closely related (as demonstrated in this literature review by the comparison of the views of boundary crossing key authors), management and organizational science views about boundary objects can be further divided based on their main emphases; management is largely focused on the use of space and collective activities involving compromise, while organizational science outlines specific criteria for events that involve boundary crossing as well as the sharing of knowledge and the efficiency with which knowledge is shared.

The views of boundary objects in management research reflect much of the French philosophy and sociology from the 1960's onward. In this school, boundary objects are referred to as "mediation objects" and serve as (largely) physical intermediary objects between respective actors. Particularly influential were Baudrillard (1968) and Foucault (1977). Both Baudrillard and Foucault discussed the arrangement of spaces to influence function; Baudrillard (1968) with regard to conveying social status through the function of furniture, and Foucault (1977) through the design of architecture with the end goal of monitoring prisoners. Hussenot and Missonier (2010) furthered this line of research centered on physical objects. They sought to understand how individuals identified the role and nature of mediation objects, their roles in collective activities, and how the

objects and their meanings evolved within their respective organizational processes.

Researchers in organizational science view boundary objects in four different ways: 1) as physical objects, 2) as ideas that can be tailored to particular needs, 3) the sharing of knowledge and conflicting knowledge, and 4) as catalysts for innovation. Major points of discussion in this section come from organizational science, organizational learning, and other journals related to organizations. A common unit of analysis in organizational science is the "organization", which can be defined as (Lawrence and Lorsch 1967, 3):

> A system of interrelated behaviors of people who are performing a task that has been differentiated into several distinct subsystems, and each subsystem performing a portion of the task, and the efforts of each being integrated to achieve effective performance of the system.

Distinctions between types of objects are made on various axes of "differentiation" and "integration". Differentiation, when it pertains to the organization itself is most simply defined as the "state of segmentation of the organizational system into subsystems" (Lawrence and Lorsch 1967, 3-4). Differentiation can be analyzed by examining each subsystem's functions in communicating knowledge between communities: the physical or tangible versus the abstract; or the objects' functions for transference, translation, and transformation of knowledge (Lawrence and Lorsch 1967, Carlile 2004). Integration is "the processes of achieving unity of effort among the various subsystems in the accomplishment of the organization's task" (Lawrence and Lorsch 1967, 4). Organizational function performance is dependent on the types of relationships between differentiation and integration (Lawrence and Lorsch, 10). Namely, the more roles that are created for accomplishing tasks, the more difficult it is to achieve integration.

Management research often relies on the methodology of actor-network theory to explore relationships between boundary objects; namely, the analysis of actors in relation to one another and said objects. Hussenot and Missonier (2010) sought to understand how individuals identified both the role of, and nature of, mediation objects using actor-network theory, focusing on actors' interactions with mediating objects when engaged in collective activity. Additionally, they explored their roles in collective activities, and how

the objects and their respective meanings evolved within distinct organizational processes. Jeantet and Vinck (1995) describe the mediation object as the result of compromise regarding a matter of communication, requiring compromise between actors involved with the object. Jeantet and Vinck viewed boundary objects as fulfilling one of three roles: compromise, prescription, or controversy.

Flichy (2007) discussed boundary objects as events during which the potential for growth and development of translation/understanding between entities. In her view, a "catchall object," or brief-lived event in which actors must confront their own views regarding an idea or technology, calling it an "ideological balloon which will subsequently be deflated when the promoters of the new technology are faced with the hard realities of production and marketing" (Flichy 2007, 159).

Carlile (2004) and Howard-Grenville and Carlile (2006) described the role of boundary objects to resolve consequences arising from the interdependency of different kinds of knowledge. Specifically, Howard-Grenville and Carlile discussed boundary objects as they are used to address conflicts arising from the transplantation of knowledge from one area to a novel application in another. Their discussion centered around the creation of points of translation between ideas and regimes previously deemed incompatible in order to achieve congruence.

Sensemaking, or the development of understanding derived from a situation, can be a form of boundary crossing. Weick (1988) discussion of sensemaking focuses on the cognitive activity of individuals as they develop insight into situations and thus, find meaning. For Weick, boundary objects were pragmatic. Holford et al. (2008) recommended viewing boundary objects as boundary "constructions," as successful interchange and translation of knowledge creates constructs between boundaries, and must be carefully constructed to ensure the successful sharing and mediation of knowledge between actors in diverse communities. Sensemaking, with regard to project development, focuses on the process of development rather than the end product (Papadimitriou and Pellegrin 2007, Williams 2005).

The usefulness of boundary objects as tools in organizational science remains de-

bated, as their inherent ambiguity can make them difficult to update when understandings change. From the implementation of the boundary object as a theoretical device ranging to an existing explicit entity, "their (in)ability to structure meaning and action: objects can be loose or ambiguous, more or less strongly structured depending on the situation" (Papadimitriou and Pellegrin 2007, 438). In other words, Papadimitriou and Pellegrin argue that the usefulness of a boundary object is situationally dependent, and may or may not provide sufficient structure to guide the translation of knowledge.

Closely related to organizational science and management, information studies focuses on technological gatekeepers and their roles in innovation. Such innovations range from the use of new sociotechnical constructs and categories, to innovations in surgical sterility. The use of new constructs as sociotechnical categories that have multiple functions and identities depends on the groups that use them, as in the case of a new forest category in forest management (Akerman et al. 2010).

2.2.6 Information Science

Information science encompasses several sub-specialties. In general, boundary object research is focused on ways to enhance knowledge sharing and examine acts of boundary crossing. Boundaries are seen as mutable constructs, but traditions and social restrictions can result in certain crossings being considered as taboo (Sugimoto and Weingart 2015). Huvila et al. (2016) reviewed literature related to the notion of the boundary object in information science, and found that the domain of information science tends to focus on the examination of boundary activities, the exploration of boundaries, and the investigation of other boundary "things" (1809). They categorized an overview of some identified types of boundaries into boundary-related activities, boundary related things, and different types of boundaries. They note that despite the implied stability of the term "object," the term boundary object is frequently both too rigid and yet unspecific to apply to many cases investigating or describing the notion in practice (Huvila et al. 2016, 7).

Boundary crossing acts in information science include borrowing theories from other

disciplines, citing across disciplinary boundaries, publishing in extra-disciplinary journals, and collaborating with members of other disciplines. Hall (2003) suggests that the tradition of borrowing theories is a hallmark of information science. This is evidenced through the adoption and integration of actor-network theory, and others. Borrowing terminology to illustrate the conceptual arc of nuances in understanding also serves as a means of crossing boundaries Ridenour (2015). Publication outside of one's parent discipline can be seen as an act of boundary crossing (Pierce 1999, Larivière et al. 2012). In 2011 alone, 60% of authors who published in LIS journals crossed boundaries by publishing in the journals of other disciplines (Larivière et al. 2012).

Boundary objects in the form of ontologies, schemas, and classifications are frequently created in response to a need to increase the interoperability of systems. Sampalli et al. (2010) used SNOMED CT, a comprehensive guide to clinical terminology, to create an ontology to enhance communication between clinicians working together to manage complex chronic conditions. The ontology was created with the input from clinicians as they re-coded a selection of mis-coded medical charts using standardized codes from SNOMED CT, allowing for the input of multiple experts in its creation.

"Pivot points", as discussed by Marchese and Smiraglia (2013) were terms that were used informally between members of a community when more official language existed. Pivot points serve as linguistic boundary objects, as they work to translate information across situational contexts for individuals in different roles in an organization. Such points become synonyms for terms, which, if captured, can serve to address gaps between situational and cognitive boundaries.

Star's views regarding the distributed and heterogeneous nature of science are highlighted in classification research's emphasis on the relationship between the heterogeneity of information and mechanisms that must be put in place for knowledge to be transmitted across boundaries (Albrechtsen and Jacob 1997, 1998, Ahlqvist 2008). Albrechtsen and Jacob (1998) further emphasize that empirical epistemology drives the creation of thesauri to span boundaries based on user, literary, or terminological warrant. Revisions to classification schemes are perceived to artificially segment knowledge and access to

knowledge beyond the temporal bounds of each version Tennis (2002). The question of relevance and boundaries becomes a case of creating and examining a scope of current literary warrant instead of a holistic view of a concept in its entirety.

2.3 Boundary Objects and Interdisciplinarity

The concept of the boundary object naturally extends to understanding interdisciplinarity. Creating ways to examine pivot points as they are present in interdisciplinary scientific literature, and span multiple levels of granularity and are present in multiple communities will result in methods of analysis to understand how interdisciplinary concepts are interrelated, and how to better classify these concepts for information retrieval systems.

As in my earlier discussion of on sociocultural navigation and individual students when discussing boundary objects in education, the phenomenological distinction between what is, and what is not yet, a part of an individual's framework of understanding must be considered when examining potential interdisciplinarity; the recognition of a newly encountered or re-examined idea or an idea parallel to ones' own research, and the incorporation of these ideas all create a potential obstacle course for a researcher's disciplinary identity. Mitigating new understandings and the syntheses they result in against institutionalized traditions (if present), can result in the production of knowledge that doesn't fit in any existing paradigm. New knowledge as it is introduced into an institutional paradigm creates problems, as issues of identity and having the newly created knowledge accepted and understood by a parent or otherwise pre-existing discipline. The creation of new knowledge results in eventual identity crises for established scientific paradigms, and causes emerging fields to vie for external and internal validation as science.

Conceptual devices and physical artifacts for translation and conveying meaning can be viewed as boundary objects. Boundary objects range from the physical and tangible to abstract concepts involving the crossing of individual or group intellectual boundaries and beliefs. Locating undefined areas of interest to multiple disciplines before they become the subject of study would solve several classical problems in classification, indexing, and

information retrieval. Making these shared problems of interest explicit can facilitate communication between interdisciplinary scholars who may wish to work with others to solve these problems.

Boundary objects themselves tend to be examined through a sociological lens, but both boundary objects and boundary crossing are examined both at the individual and institutional level across multiple private, personal, social, and work situations. Boundary objects fill gaps between entities to facilitate understanding, communication, negotiation or compromise, or to spark innovation; while also being used to convey meaning through their design. Such abstract conceptual boundary objects are frequently identified in interdisciplinary literature as "boundary concepts".

Philosophy, especially that of Foucault, delved into the idea of borrowing and crossing boundaries in the examination of human knowledge. In discussing the faces of human knowledge, Foucault (1970) addresses what is now discussed as boundary crossing behaviors. He addressed boundary crossing in two ways: 1) through concepts that are "introduced from another domain of knowledge," and 2) through constituent models, which are the efficiency creating operations. Classification and other various interpretations of information retrieval map onto these ideas, as concepts are broken down and rendered into mapping of likeness in order to facilitate efficient navigation by users within an information system.

Specifically in examining work on interdisciplinary discourse, Klein (1990) examined the act of borrowing. Her discussion echoes that of Shannon and Weaver, in that she argues that "the quality of borrowing depends upon the quality of both disciplinary and interdisciplinary communication," and the "reciprocity of 'text' and the translator" (94). This parallels work published in philosophy of science (specifically Collins (1998) and knowledge organization (Dahlberg 2006); both Collins and Dahlberg assert that the language convey meaning through time, while Collins also states that the grammar is required in order for the language to be interperable. In examining the act of borrowing in archaeology, Klein highlights that Gumerman and Phillips (1978) noted two different types of borrowing from other academic disciplines: 1) borrowing results, and 2) the

borrowing of conceptual models of causation and the borrowing of concepts.

Boundary crossing can be viewed in terms of borrowing ideas, as evidenced by scholars who adopt terms, ideas, and methodology from one research area and actively seek to integrate the terms in transdisciplinary studies within another. Boundary concepts have the power to bring together individuals to solve problems irrespective of their occupational identities, as these concepts either are related to or represent problems that span multiple fields of research. All types of named boundary entities and actions serve as conceptual surrogates to fill some need that is not met by the structure of a current system.

2.4 Measures of Interdisciplinarity

The most basic measure of the interdisciplinarity of a work is the number of Essential Science Indicators (ESI) categories cited by the work Moed (2015). Bradford scatter metrics designed to identify the level, or concentration, of interdisciplinarity of scientific fields has been attempted based on the diversity of citations, measures of concentration, and official reports. These efforts include:

- Shannon diversity index (Shannon-Wiener Index)– measures how interdisciplinary a text is based on the spread of its citations. Originally used to quantify entropy in text (Shannon 1948)
- Herfindhal index – statistical measure of concentration in economics Rhoades (1993)
- Journal Citation Reports (JCR) and Essential Science Indicators from the Web of Science (Garfield 1991, Thomson Reuters 2013)
- Leydesdorff's **?** measure of interdisciplinarity – a calculation accounting for variety, balance, and disparity
- Porter and Rafols measure of interdisciplinarity – integration calculation of Web of Science Subject Categories and journals cited by an article (Porter and Rafols 2009)

These measurements can be used to determine the reach of individual documents, and could be used to detect areas of emerging interdisciplinarity in order to effectively index literature of interest to researchers in multiple disciplines.

2.4.1 Classification and Indexing: Respecting Epistemological Stances

Classification and indexing are two closely related topics. Classification is the process of determining where concepts belong in relation to one another, where indexing is the process of assigning terms to a document. Classification provides an understanding of how a community views the world (Bowker and Star 1999), and must be updated to reflect sensitivity to cultural change Smiraglia (2014)–that is to say, if a classification is not updated to reflect community views, it loses its relevance. Domain-specific lenses are often invisible for information in bibliographic databases; an exception being domain-specific database thesauri present in EBSCO products, including LISTA and PsychInfo. Choices made in classification reflect the representation of epistemological stances of disciplines, and these choices can affect interoperability of knowledge organization systems when applied to more than one domain. Automatic classification relies on the pre-processing of text according to certain prescribed methods by an authority on the subject. Many of these work to group documents into clusters based on statistical features of the text.

Indexing is the process of assigning terms from a controlled vocabulary to a document. Humanistic approaches to indexing examine the content and meaning of text, as opposed to the statistical likelihood that a group of texts or an individual term is related to another text or term. Human indexers engage in subject analysis of both the document and expression prior to assigning terms from a controlled vocabulary. This analysis requires domain knowledge, and is both expensive and time consuming. Different approaches to indexing influence the outcome, "aboutness" versus examining the concept of "subject" produce different results; as such, subject analysis requires inferring associated meanings beyond the context of the text itself (Albrechtsen 1993). These approaches can be aided and informed by domain analysis, which examines more than just text; domain-analytic paradigms seek to investigate a community to reveal the underlying teleology, shared hypotheses, epistemological agreements, methodologies, and social semantics (Smiraglia and Lee 2012). Indexing documents that are interdisciplinary in nature may require consideration of isolated concepts; information about the social semantics of the community. Statistical methods calculate the likelihood that one term is related to another. These

can be used to automatically create thesauri; however, to be valid, the sample collection of documents must be similar enough to the target domain of the thesaurus in order to be valid (Salton 1989).

Statistical models for classification were developed to address the need to process and classify large amounts of text. Statistical and computational linguistic techniques for indexing use frequencies of term occurrences that can be used to automatically create thesauri. In order for the thesaurus classes to be valid, the collection of sample documents must be similar in nature to the thesaurus application area Salton (1989). These term matches can be extended by removing suffixes and conducting matching on word stems against dictionaries, thus creating more potential matches Salton (1989).

Classifying interdisciplinary documents can be problematic. In the current journal-analytic paradigm of citation databases, work is classified according to its venue of publication and not its conceptual content. In the case of systems such as WoS and Scopus, additional keywords are added through indexing to describe the content of each document. Pictures can be used as a thought experiment surrogate when thinking about interdisciplinary documents. Foucault (1970) discussed the multiplicity of perspectives contained in the painting, Las Meninas, in the first chapter of The order of things: An archaeology of the human sciences. In considering interdisciplinarity and aboutness, the diverse perspectives that contribute to the creation of, or the describing of, any document must be considered in order for the document to be accessible by as large an audience as possible. Classification should accommodate for this diversity of perspectives through enumeration of the objects and the points of view to be reflected (Otlet 1990, van den Huevel and Smiraglia 2013).

2.4.2 Measuring Indexing Terms: Exhaustivity and Specificity

In indexing, it has been found that the assignment of terms occurring in the mid-frequency range of occurrence are best for retrieval, that is individual terms occurring with a mid-range frequency of occurrence (Salton 1975, Wolfram 2003). Exhaustivity and specificity are statistical properties of indexing terms; exhaustivity is a measure that describes the

number of terms assigned to a document, and specificity describes the level of granularity of term assignment to a document (Sparck Jones 1972, Salton 1975, Wolfram 2003). Should an interdisciplinary document be indexed using current classification practices, it is possible that a large number of terms would be assigned to the document in order to insure that it is retrieved. Levels of granularity can differ from discipline to discipline, and incorporating unique perspectives into a controlled vocabulary requires a great deal of time and effort. Disciplines that are still growing and are interdisciplinary in nature pose exceptional problems in indexing, as science is classified according to prescriptive categories before the science is done.

2.4.3 Classification of Journals

The entirety of the works published in a journal are assigned only one, and only one, classificatory space. This has been advocated for in an ex ante manner, meaning the classification is created before the science itself is done (Leydesdorff 2006). This type of classification requires that new journals and fields be either excluded or categorized according to the existing system (Leydesdorff 2006). As classification is established prior to the creation of knowledge and literature in the sciences, the ex ante framework allows for the evaluation of science as it progresses and evolves.

When examining the interrelatedness of journals, Narin et al. (1972) used citing relationships between journals contrasted against each journals' respective classifications to examine the "mosaic" of science presented by their dataset. Their study examined the seemingly intuitive sequence of the flow of citation from mathematics to biology, inspecting citation relationships to determine boundaries between disciplines and cross-disciplinary interactions. Determining these boundaries can be an involved political process, as different conceptual camps can form in reaction to shifts in understandings of science.

2.5 Conclusion

In conclusion, boundary objects are tangible or abstract entities that inhabit shared spaces, and boundary concepts are those concepts that transcend the bounds of a discipline. Concepts contained within documents are identified and given surrogates through the act of indexing, and documents themselves are placed in conceptually similar spaces based on the examination of work done in the past. Boundaries themselves can be arbitrarily based on presumed meaning through computation of existing works, or created and enforced by actors within living communities. The types of boundary separating entities or social groups are classified in various ways, and their respective classifications provide a lens through which to examine various factors that may inhibit, or conversely, help, cross-disciplinary interaction. Disciplines that discuss boundary objects are largely (but not exclusively) social sciences, and include sociology, medicine, education and psychology, management and organizational science, information systems, and information science. Studies that examine boundary-inhabiting and -crossing entities include those involving physical artifacts (Star and Greisemer 1989), crossing disciplinary boundaries through citation (Pierce 1999; Larivière, Sugimoto, and Cronin 2012), linguistic boundary objects in a workplace (Marchese and Smiraglia 2013), ontologies as boundary objects (Sampalli et al. 2010), and various applications in information science (Huvila et al. 2016). Although studies such as Sampalli et al. (2010) and Marchese and Smiraglia (2013) examine boundary objects as having linguistic function, these studies were conducted using qualitative methods in the context of a workplace and not as text inside an information system.

In all perspectives discussed in this chapter, boundary objects as a topic share the property of being a conceptual device for making sense of an entity of interest to multiple communities of practice. Classifications, whether human-created or automatic, create representations of the relationships between communities (in the case of journals) and the concepts which they contain. These communities of practice align to domains, or sub-domains, in the realm of domain analysis. Examining the interrelatedness of domains and sub-domains of interdisciplinary nature in turn provides a means by which to

understand the connections between communities.

Chapter 3

Review of Methods

3 Review of Methods

3.1 Text Processing and Informetrics

Information retrieval and informetrics can be used to inform the designs of one another. Informetric characteristics of databases, such as term frequencies, can be used to determine optimal structures for efficient retrieval (Wolfram 1992). The informetric characteristics of databases can illuminate underlying trends within the database such as the distribution of index terms, term exhaustivity, term co-occurrence, document citation, document co-citations, and database growth Wolfram (2000). These characteristics can be used to analyze trends within the data over time, illuminating historical patterns of disciplinary and interdisciplinary collaboration.

3.2 Citation Analysis

Citation analysis is used to reveal underlying relationships between entities. Distinctions can be made between references and citations; references serve as an acknowledgement of the material taken from a document (a document refers to another document, which points backward in time), while citations are given (a document is cited by, which points forward in time) Egghe and Rousseau (1990). Citation implies a positive instance of linkage and is seen to validate claims made by other documents through their inclusion in a bibliography, but a lack of citation may not imply the opposite; that is to say that a

lack of citation of what one scholar deems to be a relevant document may not be accepted as a valid source by their community of practice.

Unfortunately, Cronin's (1981) call for a theory of citation has yet to be satisfied. Although attempts have been made to explain citation behavior, few qualify as a theory. Citations alone can be used for mapping flows of scientific information, but scientometric methods can be extended to analysis of the textual content of documents. Titles, terms, abstracts, keywords, and full text can be used as a source of data for richer topical analysis, including detecting conceptual and topical associations (De Bellis 2009).

Practical sources for gathering citation data include the Web of Science and its databases, Scopus, and Ebsco products. In the following sections I discuss the types of citation analysis and the types of phenomena to which they are applied.

3.3 Mapping Citation Flows

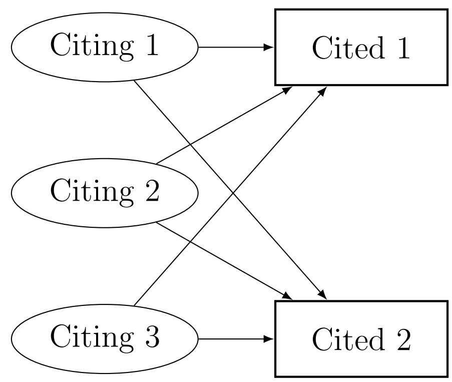
Citation analysis shows the flow of how members of communities cite one another, and can be used to trace interdisciplinary communication. Pieters and Baumgartner 2002 traced the citation flows found in 42 influential economics journals between 1995 and 1997, as mapped to their parent industrial organizations. Overwhelmingly, they found that clusters of disciplines cited journals nearly unidirectionally, such as finance almost exclusively citing economics, and business and social science fields including psychology and sociology were found to be cited by and cite economics journals but not the other way around. The exceptions were the clusters of methodology and general interest journals that had reciprocal citation relationships with other journal clusters.

3.4 Co-citation Analysis

Co-citation can be analyzed using various units of analysis—including authors, documents, journals, and institutions—in a variety of combinations. In general, co-citation examines clusters of entities grouped together by a citing source; their measure of relatedness is the number of times clusters are cited together, as the more times entities are cited together, the more closely they are considered to be related. Co-citation analysis

seeks to reveal the intellectual structure of science, or the structures found in a particular area of science; in all of co-citation analysis, entities cited together are considered to be related to one another in some way by the citing entities.

Figure 3.1: Co-citation Analysis

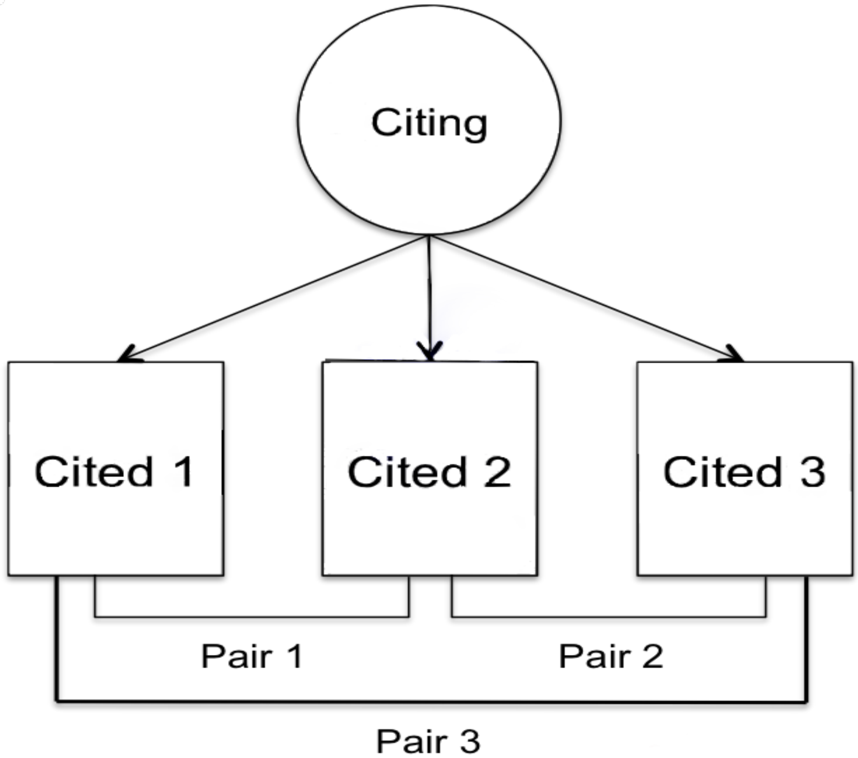


In document co-citation analysis, papers are determined to be more strongly related to one another the more times they are cited together Small (1973). Papers that are frequently cited together are considered to make up the core of the body of literature being examined. Calculating the proximity of co-citation reference pairs based off of their positions in the full text of citing documents can improve accuracy Boyack et al. (2013).

3.5 Bibliographic Coupling

Bibliographic coupling, pioneered by Kessler in the 1960's, examines coupling patterns found in the references of scientific literature Kessler (1963). As in other types of citation analysis, the strength of the coupling of a pair of items is assumed to be a measure of the citing source's assertion of the items' similarity based on the frequency of co-occurrences Ahlgren and Jarneving (2008). Bibliographic coupling can be thought of as comparing reference lists to detect co-occurrences. When compared against the Science Citation Index, bibliographic coupling can also reveal information about the intellectual structure of a field.

Figure 3.2: Bibliographic Coupling



3.6 Author Co-citation Analysis

Author co-citation analysis (ACA) can be seen as the inverse of bibliographic coupling. Where bibliographic coupling points to the intellectual structures of the past, ACA points to the future. Author co-citation analysis can be used to determine the intellectual structure of a field based on authors who are highly cited in that field against either documents or journals. Documents and journals can be classified into their respective fields, and then compared against individual authors as a measure of an author's disciplinary reach—the more fields associated with an author, the wider their reach. Authors' influences and authoring relationships can be determined using this type of analysis, and by extension, this analysis can demonstrate judgments made regarding the similarity and subject matter by those who cite them White and McCain (1998). As data regarding the affiliations of the authors to institutions or fields is included, this type of analysis can contain richer data than other types of co-occurrence networks.

Visualizing ACA requires a dataset consisting of many works over time broken down

into discrete time periods. The choice of which authors to map in this type of analysis will impact the visualization of the author co-citation relationships within domain, as different pairs of authors may be perceived to have different degrees of similarity by those who cite them. As such, choosing authors considered to be key to the area being analyzed is critical to creating a map that resembles the dominant cognitive structure of the field or fields in question.

3.7 Co-word Analysis

Co-word analysis can be used to reconcile multiple outputs of knowledge and the perspectives contributed by identifying common patterns based on word association strength in a heterogeneous information environment (De Bellis 2009, Ding et al. 2001). Co-word analysis is frequently used to map the "intellectual structure" of areas of science by identifying words and phrases with high frequencies of co-occurrence. Patterns revealed from clusters of co-occurring keywords can reveal underlying themes, and can be used as a basis on which to map science. This type of analysis can be conducted at various levels of the text, including titles, abstracts, full-text, and keywords. In an information retrieval environment, co-word analyses of keywords have been used to reduce large numbers of descriptors to multiple smaller spaces, allowing for the relatedness of terms to be more easily identified, understood, and extracted as patterns (Ding, Chowdhury, and Foo 2001). More recent studies using co-word analysis, including Uyar et al. (2020) and Corrales-Garay et al. (2019) also used temporal slices, using the most recent low-frequency co-occurrences to indicate promising areas of potential future research.

Keyword-based co-words analyses can be subset into the following: author-provided (Courtial 1994), database-assigned, expert-suggested (Looze and Lamarie), or non-parametric (Ravikumar and Singh 2015). The intellectual structure is revealed in a network of co-occurring terms and allows for the creation and examination of bridges created by scientists who are "manipulating and adapting links that are already in place," even though these links are not stable because the passage of time alters the relationships between concepts and problems (Courtial and Law 1989, 301). The unstable nature of co-word

links is demonstrated when comparing studies examining less-established areas of interdisciplinarity to those that are more established. In an early study, Law et al. (1988) used co-word analysis of keywords to detect and illustrate major research themes related to acid rain research in multiple disciplines from papers retrieved in a PASCAL database on acidification research. Co-occurring keywords were grouped and frequencies per paper were counted, and links between similar and different themes of research were identified and mapped according to their Science Citation Index indicators. Their results in acidification lacked a cohesive thematic core, as opposed to Courtial and Law's (1989) study on artificial intelligence, which found a cohesive thematic core. Co-word analysis linked overarching themes of speech recognition and character recognition to the terms associated with describing these ideas.

Recently, co-word analysis has been used to examine literature to detect "knowledge areas" in which open data is examined, the conceptual structure of open data research, and to propose trends in future open data research Corrales-Garay et al. (2019). Knowledge areas, though not explicitly defined by the authors, align to what information science would consider a "discipline," as they state that "research on open data is multidisciplinary" (Corrales-Garay et al. 2019, 78). The results of this project indicated the recent nature of the topic of open data, and that open data is of interest to a breadth of knowledge areas. To select a population of papers, Corrales-Garay et al. (2019) created a search protocol in the Web of Science using the search terms "open data" for all available years (up to the year 2017), and filtered to include only those works that themselves used open data in order for it to be reproduceable. SciMAT was used to map themes in the data. Frequency analysis using the unit of the article showed a growth trend that became exponential after the year 2012. Identified keywords were grouped (e.g. "app" and "application") and standardized to a singular form. Terms were further grouped, and statistical techniques used by works in the data sample were eliminated. Terms considered to be too general to bear meaning, such as single word terms like "information," were also eliminated. The resulting keyword sample contained 4,981 keywords. Keywords were mapped to create subnetworks of co-occurrences in each knowledge area, but were

not used to examine how individual concepts aligned across all areas.

Gan et al. (2019) used co-word analysis of MeSh terms for a pre-determined set of retrieval terms related to epilepsy genetics published in PubMed between 2009 and 2018. This paper used bibliometrics, co-word analysis of MeSH terms (keywords), social network analysis, geographic analysis of publication origins to analyze the research front of epilepsy genetics. In this work, the authors focused on the five largest clusters created by the default VosViewer algorithm. All potential papers were reviewed by two independent experts, who reached concensus on all items selected for inclusion. Bibliographic Item Co-Occurrence Matrix Builder (BICOMB) software was used to build a co-occurrence matrix, for selected articles selected for inclusion, and graphs were created using both Excel and ArcGIS. VosViewer was used to separately map the social network of authors and co-occurring MeSH terms. Both unique high-frequency MeSH terms and a discussion of each of five clusters on specific topics were included, as well as a discussion of overlap in MeSH terminology between cluster 1 and cluster 2. This work represents a combination approach to examining the research fronts in epilepsy genetics using co-occurrences in a comprehensive controlled vocabulary.

Uyar et al. (2020) used co-word analysis to map the conceptual structure of financial auditing using author-supplied keywords from 22 accounting journals indexed by the Web of Science's Social Science Citation index, plus four additional auditing journals not included in the WoS, for the years 2000-2016 to five year increments of 2000-2016; as well as to map concentrations of co-occurrences using betweenness centrality and degree centrality for continental regions, and for the entirety of selected data; and additionally, to determine the specificity of the nature of auditing keywords adopted by authors. Uyar et al. (2020) did not discuss the average number of author-assigned keywords, or whether any articles were omitted for not including author-supplied keywords. They found that over time the most general keyword, "auditing," decreased in its adoption by authors and was replaced by more specific terms related to auditing, such as "auditing fees." Regional analysis showed convergence for higher-ranked terms indicating the core of intellectual work done in auditing, and deviation in focus to region-specific interests at lower-ranking

terms. Results also indicated potential future research for what the Uyar et al. (2020) termed "promising and unsersearched themes" [69], including "continuous auditing" and "big data."

Tijssen and Van Raan (1994) implemented co-classification analysis of Dutch Research and Development (R&D) subject headings to analyze the interdisciplinary structure of research and development in bibliometric data in the database "Energy Science & Technology" for two three year periods (1984-1986 and 1987-1989). Co-classified data was examined for co-occurring classifications within a system of forty possible primary classifications, and numerous secondary and tertiary classifications mostly related to sub-fields and research areas subordinate to the primary areas. The analysis of co-occurrences in time slices allowed for the examination of changes in the intellectual structure of the field.

3.8 Topic Delineation and Community Detection

Delineating topics, subjects, and fields has been of interest to researchers in information retrieval for many years. In topic delineation, researchers create processes and algorithms to detect communities; communities can be siloed from one another, or can overlap. Delineation is aided by multiple information perspectives, that is, multiple representations of topics provide a way to converge on objects examined Zitt (2015). Micro-level topic delineation is an issue frequently examined in bibliometric studies; however, meso-level delineation is a problem rarely addressed (Zitt 2015). In the process of field delineation, the granularity of the field examined affects recall and precision. In an a priori system, the recall of a global query is affected by the classification and granularity of the topic and views held of the topic by experts reviewing the system, as controversial, interdisciplinary, and multidisciplinary topics exist within domains of which the borders are disputed. Therefore, the representation in the top-down system may be influenced by nuances of the expert's specialty in their review of the system (Zitt 2015, 2227).

Cabanac et al. (2020) described three "models" and one hybrid approach to identifying targeted domains and two levels of scope, or granularity, of analysis. Zitt clarifies definitions for two widely accepted scale levels fundamental to bibliometric analysis, the

meso-level, and the micro-level. The meso-level encompasses " sub-disciplines, fields, large research areas" [129]. However, he notes that "the frontier with the micro-level of research fronts or topics is quite fuzzy" [129]. He describes field delineation as being understood in types of publication sets, and the requirements for accomplishing field delineation vary from basic to demanding in relation to the domain in question. Emerging, transversal, or complex domains are "difficult" to delineate. All three models draw on quantitative methods. Zitt's "Model A" uses static knowledge, and broadly fits areas of top-down control. These include areas relying on some form of classification; he stresses that distinctions between national vs. international, institutional, and stakeholder-driven classifications are political, and not neutral, or value-agnostic, in nature. "Model B" relies on the knowledge of experts, and consists of categories made by experts' understanding of particular applications, such as patents, or journal-based citations in a given citation database. "Model C" encompasses bibliometric mapping and clustering techniques that identify hidden structures including actors, texts, and citations.

Community detection is a graph-based method of identifying communities through topical analysis, topological analysis, or a combination thereof, which can be used for classification of documents. Community detection can be an outcome of any one or combination of the aforementioned types of analyses in which clusters of co-authorship, cited papers, authors, journals, or groups of co-occurring words are examined Ding (2011). Methods to detect communities can be hybridizations of or multimethods involving both text processing methods, such as latent dirichlet allocation, and bibliometrics. Detecting communities of shared interest can help identify communities and their relations to one another. Classification of documents is done by separating distinct clusters along their respective vertices Fortunato (2010). Hierarchical relationships found in the graph can be indicative of real-world relationships, as smaller communities of researchers frequently inhabit larger communities, such as institutions Fortunato (2010).

3.9 Applications of Text Processing to Interdisciplinarity

Text processing, as applied to texts from interdisciplinary work, frequently combines multiple methods to create a clearer picture of what is going on in the field. This section discusses text processing as it has been applied to interdisciplinary work, specifically Latent Semantic Indexing (LSI), Latent Dirichlet Allocation (LDA), bibliometric or informetric approaches to text processing, and support vector machines.

Studies implementing LSI examining interdisciplinary fields use statistically derived factors to classify documents. Janssens et al. (2006) mapped full-text papers and notes published in five LIS journals between 2002 and 2004 using multidimensional scaling (MDS). Text was extracted from PDFs and DOCs and was processed to strip references, acknowledgements, author names, stop words, template words, URLs, and email addresses. Phrases and synonyms were detected, and were added back when determined to be a necessary part of a noun phrase using LT POS and LT CHUNK (Edinburgh Language Technology Group 2015), a part-of-speech tagger and chunker that uses Hidden Markov Models (HMM) for disambiguation. A HMM is a type of statistical Markov model that is assumed to be a Markov process, or a type of decision process in which actions or states transition from one to another based on a computational model, in a "hidden" state. This means that the state is not directly observable. Silhouette values were determined for documents; the Silhouette value is a statistically derived factor for a document ranging from -1 to +1, and was used to determine the best placement for a document in a cluster. This was done for numbers of clusters ranging from 2 to 25; six was determined to have the best fit, so was used as the basis for the clustering. Both "hard" k-means and hierarchical clustering algorithms were tested, which allowed for each document to be a member of only one of the six clusters. A dendrogram was used to visualize the clusters, and the cutoff at the sixth cluster. Clustering examined in this paper can also be viewed as a type of bibliographic coupling, making this a hybrid approach. Results of the LSI term by document matrix were visualized using MDS, but the tf-idf matrix for the clusters showed there was potential for overlap not accounted for by hard clustering.

Paul and Girju (2009) used LDA and their own classification algorithm to classify papers and detect trends in topicality based on each paper's topics and languages discussed in each paper. Data were gathered from the ACL anthology and included linguistics, computational linguistics, and education journals. Titles and abstracts were used for each research area, with the addition of full text when it was available. Their method sought to overcome a well-known shortcoming in citation analysis—fewer than 90% of scientific papers published are ever cited, omitting a great deal of literature from the analysis of topicality when done using citations. Some of the papers in each dataset had been pre-categorized by the publishers, but those that were not were labeled manually. Instead of relying on a binary classifier to determine category membership, papers were allowed to belong to multiple classes and were assigned multiple classes. This manual coding was done on the premise that the omission of a class label did not imply a negative class instance; a binary classifier would have interpreted the absence of a label as a negative class instance. Topics within each field were found to overlap based on the results of the classifier, as well as topics across fields. To compare interdisciplinary topics, a meta-document of each topic was created containing concatenations of words in each class; each topic was then represented as a vector of words in each document weighted by their tf-idf values, and each value was compared by similarity computation of their respective cosines. Relations between each field were visualized as four tables of the three fields, networking related terms by drawing weighted lines between the term tables created for each discipline (Paul and Girju 2009, 241).

3.10 Metrics-based Approaches to Text Processing and Information Retrieval

Both knowledge organization and information retrieval implement top-down, or a priori, approach; knowledge organization creates classifications that are employed by information retrieval in order to create what Zitt (2015) describes as "structured expectations" (2226) on which to base formulas for good retrieval performance. Bibliometric approaches, as opposed to IR approaches, present an a posteriori, or bottom-up approach, most similar

to natural language processing in that it relies on clusters that emerge from data in a dataset. Scientometrics tends to serve a broader audience than information retrieval, seeking to inform research and policy where information retrieval seeks to fulfill the needs of the user Mayr and Scharnhorst (2015).

Bibliometrics, informetrics, and other metrics approaches can be used to inform information retrieval. Bridging traditional top-down IR techniques with bottom-up bibliometric techniques can provide a solution to classification bias and interdisciplinary problems (Zitt 2015). Parallel relationships between entities studied in informetrics and information retrieval are apparent—both use a dual, or symmetric, relationship to identify and examine entities in question, such as citing and cited in bibliometrics, and documents and queries in information retrieval (Wolfram 2015, Egghe 1990). Network-based queries can be mapped, and ideally, this type of mixed approach can be used to solve issues found in information retrieval systems Zitt and Bassecoulard (2006).

Lexical queries, sometimes also called bibliometric queries, are used to create datasets for bibliometric analysis (Maghrebi et al. 2011, Zitt and Bassecoulard 2006). These queries are compound, consisting of sets of keywords that are logically organized with appropriate Boolean operators in order to retrieve a set of papers that fall within the defined scope of the research area or discipline to be examined (Maghrebi et al. 2011, Zitt and Bassecoulard 2006).

Much of informetrics is aimed at mapping science and how disciplines relate to and communicate with one another. In order to understand this, a “seed” or “core” of literature is used to create an extended dataset; this is done by adding “conditional criteria” to an initial query which expand the set of documents that may be returned to a query, thus extending the returned set of potentially relevant documents (Zitt and Bassecoulard 2006, Glänzel 2015). Documents are considered relevant if they meet certain conditions related to core documents, but expert analysis may be needed to determine relevant from not-relevant documents, as such distinctions may be “fuzzy” (Glänzel 2015, 2217). Software has been developed to quantify, analyze, and visualize relations found in bibliometric and other types of networks. This software can be used in conjunction with text processing

methods to confirm and expand upon classified results of text processing.

3.11 Hybridization of Methods

Hybrid methods frequently combine a text processing method with a bibliometric method in order to create a complete picture through triangulation. Such hybridized methods examine heterogeneous data. For example, Janssens et al. (2006) implemented a hybrid approach that integrated citation analysis and textual analysis. This was done using a term-by-document matrix to examine citations in combination with a cited references-by-document matrix.

3.12 Support Vector Machines and Bibliometrics

Though bibliometric in origin, Borgman's 1989 proposed framework for examining the intersection of variables studied and the research questions asked influenced investigations into interdisciplinarity using support vector machines. Demarest and Sugimoto (2015) added the object of discourse to Borgman's framework, proposing the study of "discourse epistemetrics". In their pilot study implementing the proposed method of discourse epistemetrics, they applied an SVM to a sample of doctoral dissertations from ProQuest for the mutually exclusive topics psychology, physics, and philosophy. Terms identified to discriminate between disciplines were assessed according to frequencies of occurrence, and included words such as "calculate," "learn," "observe," "assessed," and "used." Demarest and Sugimoto (2015) suggest that these types of words show underlying social and epistemic differences between the disciplines; I would add that this is apart from the topics considered core to each discipline. The authors do note that term frequencies neglect underlying semantics, which could potentially be addressed by expanding the window of analysis.

3.13 Author Topic Modeling

Sugimoto et al. (2011) used LDA to map the changes in topicality of 3,121 library and information science doctoral dissertations published between 1930 and 2009. Data were

gathered from the MPACT database, which contains doctoral dissertations; those that were not in the dataset as full-text were retrieved and added to the database as such. The study analyzed the full text of each dissertation. Data were sliced into separate time periods to analyze the identified core areas in library and information science. It was found that these core areas did not change substantially in topicality in any period of time examined in the study.

3.14 Citation Analysis for the Examination of Interdisciplinarity

Citation analysis examines the underlying structures of references between homogenous and heterogeneous entities. Yan and Ding's (2012) study compared six types of citation analysis used data consisting of citations and text from 59 information science journals indexed in 2008 by Thomson Reuters. Documents gathered were published between 1965 and 2010, but only three selected time periods between 1991 and 2010 were analyzed. Data were processed to filter documents by institutions and the year of the citing papers' publication. Yan and Ding (2012) discuss and compare six types of citations found in scholarly communication, identifying two dimensions among the types of citation networks on the basis of cosine distance: "citation vs. non-citation", and "social vs. cognitive". Of the six types of citations examined, co-occurrence is the best at determining interdisciplinarity through the examination of co-authorship. Co-authorship clustering was found to be much denser than other type of clusters. Real-world social connections such as invisible colleges implied by co-authorship networks were more difficult to establish than statistical similarity approximations between documents; the majority of co-authorship was found to occur in established institutions such as universities.

Mohammadi and Thelwall (2014) examined correlations between Mendeley readership counts and citation measures in social science and humanities disciplines, and compared cross-disciplinary information flows as demonstrated by Mendeley bookmarks to citation patterns in the Web of Science. Cross-disciplinary flows were determined by measuring the percentage, as opposed to the number, of readers in one WoS category reading articles

from other categories. The change in percentages of categories per year cited by each discipline was examined.

Fang (2015) used document co-citation analysis to examine the disciplinary origins of the domain of digital medicine using CiteSpace. His dataset was gathered from the Web of Science using a bibliometric query, then processed by slicing into time frames of one year, limiting to references cited n ¿ 200 times, and pruning and merging the resulting data. Document co-citation clusters were detected and mapped using CiteSpace and visualized by displaying bursts of clusters according bursts mapped in a matrix of yearly time slices by keywords.

3.15 Visualization and Science Mapping

Efforts to examine interdisciplinary research areas to detect scientific and technological trends use bibliometric delineation of published articles, patents, and other literature. Some studies, such as the one conducted by Maghrebi et al. (2011), use an accepted definition of a field in order to set boundaries for the research area (in this case, the emerging field of nanotechnology). Keywords that were not deemed relevant based on the operating definition of the field were removed in order to build a lexical query. If precise enough, this type of query can be used to create datasets of interdisciplinary material regardless of how documents are initially indexed in a database.

Price 1965a, 1965b first described the process of science mapping in the 1960's, discussing how science mapping could convey information gathered from scientific literature using bibliographic data. Network visualization of bibliometric data can provide spatial orientation, allowing the user to identify a sense of place in the visualization akin to that of a geographical map Small (1973). Both small and large-scale efforts have been made to visualize the structure of different domains. Interdisciplinary efforts have included large maps of science itself; these relationships have been traced based on different methodological premises. Inter-journal citation relationships have been used to map interdisciplinarity Boyack et al. (2005), such relations reflect the heterogeneous affiliations of many journals (Bradford 1934, Garfield 1972, Leydesdorff and Rafols 2009).

Smaller scale efforts to analyze microcosms of interdisciplinary began in the 1980's, beginning as co-word or bibliometric-based analyses. These include Law et al. (1988) and Courtial and Law (1989). Tijssen and Van Raan (1994), Maghrebi (2011), and others followed suit. Larger scale efforts fo visualize the underlying structure of science rely on existing classification systems, such as the Clarivate ISI journal classifications (formerly Thomson Reuters). Boyack et al. (2005) developed a comprehensive map of science based on Web of Science^{TM} journals using a variety of journal inter-citation and co-citation measures. Their map included over a million records from 7,349 journals. In this case, different journal-journal similarity measures resulted in different maps, but all used the same data.

Co-classification, co-word, and visualization all fall under the penumbra of "science mapping." Law et al.'s (1988) study on acidification research was part of an early effort to create tools that anyone could use to examine the structure of science. Porter and Rafols (2009) sought to measure the interdisciplinarity of science using articles from six of the 244 Web of Science Subject Categories published in eleven year intervals. The indicators of interdisciplinarity they used included citations within and outside of Subject Categories, the number of authors per paper, the number of Subject Categories cited per paper, and an "Integration Score." They found that, on the basis of their measures, science was becoming more interdisciplinary. As more subject categories were created to accommodate growth in the sciences, naturally, more subject categories were cited following the chronological progression of time.

Tijssen (1992) examined co-occurrences of classification, terming it co-classification analysis. Tijssen's goal was to aid Dutch scientific policy in the field of energy research. Data consisted of seven years' of publications (1984-1986 and 1987-1989) contained in the U.S. Department of Energy's Energy Science & Technology database. For each publication, co-occurrences of six-digit subject headings were entered into a matrix. Indicators of interdisciplinarity were considered higher for those with higher frequencies of co-occurrence. Results of the co-occurrence matrix were plotted to create a map of the intellectual structure of energy as a field, and results of the map were verified by separate

groups of experts consisting of 1) scientists and project managers and 2) R&D program managers and science and technology policy makers. Both the first group and second groups had their reservations about the accuracy of the map, but scientists were more interested in their areas of specialty while policy makers tended to doubt the ability of bibliometric methods to aid in decision making. Tijssen concluded that quantitative features of this type of co-classification analysis can aid in identifying features of interdisciplinary and multidisciplinary areas of science, clarify linkages between specialties, provide supplemental material to traditional knowledge codification and aid in the formation of consensus between R&D program managers and research policy makers. It is important to note that though Tijssen did not describe it as such, the resulting maps of science serve as a boundary object between two distinct groups in a community of practice.

3.16 Disciplinary Diversity and Network Analysis

Disciplinary diversity affects network properties, which in turn affect network coherence. Porter et al. (2007) developed a framework of integration and specialization when examining networks for interdisciplinarity. Rafols and Meyer (2010) examined network properties for knowledge integration in interdisciplinarity, and introduce conceptual frameworks for examining interdisciplinarity within networks by examining the concepts of diversity and coherence. Diversity, a concept integral to many fields, examines the heterogeneity of network structures "in terms of variety, balance, and similarity of categories" (Rafols and Meyer 2010, 9), relies on existing top-down classificatory structures "which may miss emergent or dynamic phenomena in science" (8). The three categories of diversity are variety, balance, and disparity/similarity (Rafols and Meyer 2010; Stirling 1998, 2007). Interdisciplinary metrics can ameliorate existing multitudinous categories, but fail to address "hidden divides within existing categories" (Rafols and Meyer 2010, 8). Coherence examines the "functional articulation and structural compactness of elements in system," using a bottom-up structure independent of existing classification schemes. The four possible categories resulting from the examination of diversity and coherence of

network properties are found in Table 3.1.

Table 3.1: Rafols and Meyer's Patterns of Interdisciplinary Integration

Level	Coherence vs. Diversity
Potential interdisciplinary integration	Low coherence, high diversity
Specialized interdisciplinary	High coherence, high diversity
Potential integration within discipline	Low coherence, low diversity
Specialized disciplinary	High coherence, low diversity

3.17 Systems for Addressing Interdisciplinary Classification

While top-down, or a priori, systems may miss emerging phenomena in science (Rafols and Meyer 2010), bottom-up, or a posteriori, approaches may not update retrieval systems quickly enough to adapt to changes in fronts of science. A priori classification schemes accommodate interdisciplinary entities in one of two ways: they are either placed in an interdisciplinary category akin to an "other" category as in the case of the NARCIS classification (Smiraglia 2017), or interdisciplinary entities are attributed to a disciplinarily-bounded base. Methods to help advocate for the inclusion of interdisciplinary entities in classification schemes are developed to remedy the disciplinary-bounded nature of domain-dependent classification systems, using what is essentially footholds in each system to create ways to advocate for interdisciplinary views that have been neglected in the creation of the system. This type of approach examines interdisciplinary work from a deconstructivist standpoint.

Continuing a deconstructivist arc, Katz and Hicks (1995) and Sainte-Marie et al. (2018) both developed ways of reclassifying Web of Science data; the important distinction is that Katz and Hick's method was journal-based, and Sainte-Marie, Ridenour, and Larivière's method was paper-based. Katz and Hicks (1995) sought to address the issue of interdisciplinary classification and growth in science by examining interdisciplinary publication activity to create a schema reflecting publication and bibliometric trends present in UK interdisciplinary journals. They examined discipline-field relationships defined by categories names present in the Institute of Scientific Information (now the Web of Sci-

ence)'s 154 sub-field classifications, examining (in part) collaboration. From this, they developed a flexible 17-field classification (that can be grouped into four higher-level categories). The resulting 17-field classification scheme provided a relatively stable framework from which to analyze growth in science over time. Sainte-Marie, Ridenour, and Larivière (2018) developed a paper-based classification scheme for the transdisciplinary area of cognitive science. Their method collected papers containing the term stem "cogni*" and developed a series of reclassification algorithms to determine a better categorical home for papers using concepts present in paper titles. The resulting algorithm created an additional step toward a topically-based classification scheme dependent on the conceptual contents of the paper, and not the presumed disciplinary home or identity of the paper's publication.

3.18 Conclusion

Most text processing methods for automatic classification, including LDA/LSI and SVM, output either a linear model or a set number of classes based on the statistical fit of the selected body of text to an ideal number of clusters. Natural language processing reveals more details about the content and topicality of a curated corpus of text; however, methods of truncating text to expedite processing can result in a loss of context. These include stemming, a method that shortens a word to the most meaning-bearing unit, and lemmatization, a method that calculates the most common frequency of occurrence of a word's form and uses that particular form as the representation of the word for the entire corpus. Citation analysis methods imply positive linkages between various combinations of items including documents, authors, journals, and classes. Citations create a positivistic representation of a curated set of data that is empirical in the sense that the citation data itself is factual, and is usually bibliometric in nature.

Studies computing interdisciplinarity and disciplinary overlap in text focus on bibliometric indicators, as these are empirical in nature and are easily verified. Text processing methods can be used to examine conceptual overlap, as demonstrated by Ridenour (2015) and Sainte-Marie, Ridenour, and Larivière (2018). Text processing and bibliometrics can

be used in various hybridizations of methods to examine the content and interconnectedness of documents within a corpus, but such studies do not necessarily incorporate an examination of classificatory structures in order to examine the alignment or divergence of concepts along and across classificatory scaffolding.

Most studies examining boundary objects do so using qualitative methodologies, and many of these studies are conducted in a corporate or institution-specific environment. Though studies employing the term "boundary object" tend to follow qualitative methodology, the identification of boundary concepts as found in published literature can be done through the examination of natural language against the classificatory structures superimposed on the publications, and even the journals, in which the natural language is contained. Applying the idea of the boundary object to a database surrogate for Popper's third world where ideas interact, fills a gap in the literature by allowing for the examination of interdisciplinary interactions as artifacts in information systems, going beyond points of mutual interest within the containment of an individual institution. This allows for inter-institutional exchanges. More so, it allows for the identification of areas where researchers form different institutions who may not be aware of the other's work to locate said similar work.

Chapter 4

Methods

4 Methods

At present, no study examines the concept of the boundary object using a strict classification of journals as it exists in information systems to examine concepts. Unlike previous studies examining the notion of the boundary object in an institutional setting using qualitative methodologies, this study converges on boundary concepts as evidenced in information systems through natural language in titles and authorship patterns surrounding core concepts in scientific literature in journals indexed in the Web of Science, and by verifying resulting high-frequency occurring concepts with authors who have published about them as identified in the dataset. This is a novel approach and involves a carefully designed set of methods, of which preliminary versions have been vetted and published (Ridenour 2015 and 2016, respectively).

1) What epistemological boundaries, manifested as classificatory divisions, exist between NSF-identified disciplines that comprise cognitive science? What is the representation of work published in these? How does this change over time?

   A) How do terminological boundary objects fit against the NSF's classificatory framework? What concepts are the most interdisciplinary in the core of the domain?

   B) What boundary crossing concepts are of the highest frequency? What boundary concepts at the core-level are 100% interdisciplinary, and what are their

topical periphera? How do boundary-crossing concepts in the core change over time?

2) What are the underlying sociology (through acts of co-authorship)? What acts of author boundary crossing exist surrounding the most interdisciplinary concepts?

4.1 Data

Data consists of Web of Science data with added NSF classification codes stored in a local database, which was made available by Professor Vincent Larivière at the University of Montreal. The same dataset was used for the initial quantitative analyses, including both natural language contained in titles and coauthorship. Determination of data for inclusion was done at the journal level, and created using a seed and expand method surrounding the string "cogni*" in journal titles. This method for data selection fits a modified version of that outlined by both Zitt and Bassecoulard (2006) and Glänzel (2015) for a seed, or core, and expand model; in this case, the modification is the temporal restrictions on the conditional criteria of the extended dataset. By restricting the temporality of the works cited, or the expanded dataset, the resulting data fits a KO notion of extension, as the time frame is artificially limited for the purposes of analysis to the years 2006-2016. The first step of the data, or the core, were paper metadata from journals containing the string cogni* in the journal title. Table 4.1 ties questions to the high-level overview of how each piece of methodology was implemented.

To reduce the computation time and file size, natural language labels for higher level disciplines (Table 4.2), disciplines (Table 4.3), and core vs. extended data (Table 4.4) were recoded using numeric values. Integers have a much shorter computation time than strings, thus reducing the overall computation time for a big dataset.

Table 4.1: Research Questions and Methods

RQ	Method
1	Epistemological boundaries are imposed by journal classification in the form of a taxonomy. The classificatory boundaries serve as epistemological markers, and the categories in which journals were assigned were identified and extracted to detect the taxonomic structure using R's subsetting function and dplyr() libraries. Numbers of publications in each category by year further subset for frequencies and time-series analysis.
1.a	Titles within each category were analyzed, and then concepts that occur in each category are examined for disciplinary spread (Moed 2015), and interdisciplinary concepts occurring with the highest percentage of interdisciplinarity were selected for further analysis.
1.b	Terminology used to describe core interdisciplinary concepts were analyzed, and concepts were aligned to their respective categories over the years 2006 to 2016. Disciplinary spread is a calculation of the number of categories a boundary object spans. In this case, the number of categories that a concept inhabits.
2	Authorship habits are examined for overall trends for the domain, triangulating who works with whom in co-authorship for the most interdisciplinary concepts using network analysis.

4.2 Triangulating Boundary Concepts

Combining multiple methods creates a clear, cohesive, and accurate picture of a problem being studied. Triangulation, or the combining of methodologies, can be done in four basic ways: data triangulation, investigator triangulation, theory triangulation, and methodological triangulation Patton (1990). Triangulation uses multiple methods, as Denzin (1978, 28) noted:

> no single method ever adequately solves the problem of rival causal factors. . . .Because each method reveals different aspects of empirical reality, multiple methods of observations must be employed.

Boundary objects in the form of terminology, citation origin, and authorship, in information systems can be identified and analyzed based on the frequencies of occurrence of terms (Ridenour 2015), and further converged on using citation analysis (Ridenour

Table 4.2: High-level Discipline Recoding

GrDisc	Code
Natural Sciences and Engineering	1
Social Sciences and Humanities	2

Table 4.3: Recoded Disciplines

NSF Category	Code
Arts	1
Biology	2
Biomedical Research	3
Chemistry	4
Clinical Medicine	5
Earth and Space	6
Engineering and Technology	7
Health	8
Humanities	9
Mathematics	10
Physics	11
Professional Fields	12
Psychology	13
Social Sciences	14

Table 4.4: Data Collection Step

Data collection step	Code
Core	1
Extended	2

2016). The identification of frequencies of terminological usage in different classification-defined disciplines allows for the analysis of shared understanding as demonstrated by co-occurrences in terminology across high-level classifications. This overlap in terminology was compared over time using a time-series analysis; this was selected based on the best fit for the available data, and R's HoltzWinters() function was used for statistical smoothing. Time slices for analysis were chosen by the year, as the set of eleven years was chosen to show the flow of an increase of completed cognitive science work versus the body of work cited by core CS journals. The methods outlined here serve to extract, analyze, and align boundary concepts across multiple categories.

In Chapter 4, methods traditionally used to analyze textual corpora from both information retrieval and bibliometric methods were discussed. Support Vector Machines analyze a set of data and determine automatic classification and regression analysis, which reveals mathematical properties of the text analyzed according to goodness of fit to the model. Latent Semantic Indexing (LSI), or Latent Dirichlet Allocation (LDA) analyzes text according to the assumption that an ideal number of clusters exist, and the mixture

of topics in a set of documents reveals the nature of an individual document. LSI/LDA will omit possible topics based on the frequencies of occurrence. SVMs and LSI were not compatible with other methods used, as they result in their own classification when the goal of this analysis was to examine concepts as natural language entities in the document titles in which they occur against an already existing classification. Methods chosen were carefully considered to triangulate the underlying social structure of concepts against an epistemological framework manifested as a classification, which relied on output that was topical in nature. Methods selected were harmonious, and required the incorporation of multiple data, tools, and measures to converge on the focus Brewer and Hunter (1989), in this case, interdisciplinary concepts (exhibited as natural language entities) and the social nature of work done in a transdisciplinary area. When applied to interdisciplinary analysis to identify concepts and examine boundary crossing concepts in bibliometric data, the data source, the boundaries, and the units of analysis were first established.

First, the data source identified was Clarivate's Web of Science. Second, the boundaries used in the analysis were those of the National Science Foundation classification system, which is a taxonomy. Third, the units of analysis selected were the concept (as a noun phrase), and the published authors. Because boundaries created by classification are artificial, in that they are imposed at the journal level and do not reflect the natural social dynamics of the community, this study is most appropriately addressed as a domain analysis.

Next, the identification of methods to analyze concepts crossing boundaries was considered. Epistemological boundaries, in this case, are the silos created dividing journals along the NSF taxonomy. The combination of natural language and bibliometric methods allowed for triangulation by examining the epistemic and social dimensions of the most interdisciplinary topics. These topics met the criteria of spanning all six boundaries in the core domain of the data. Thus, these boundary concepts are boundary spanning in nature. This allowed for a clear focus to converge on transdisciplinary boundary concepts and the researchers working together on these highly interdisciplinary concepts. Social dimensions analyzed, in this case, were patterns of coauthorship surrounding identified

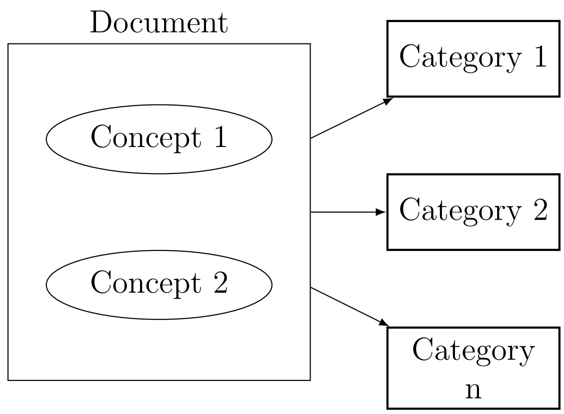
boundary spanning concepts.

4.3 Analysis of Boundary Concepts

Concepts were analyzed using publication titles in the WoS dataset. No clear advantage has been shown in the analysis of full text versus titles and abstracts. Concepts, in the case of this analysis, were based on noun phrases as calculated using TextBlob in Python. Noun phrases provided a consistent unit for the basis of analysis. The disciplinary spread of boundary concepts found in abstracts was measured by the span of each concept's interdisciplinary reach, as defined by the number of NSF categories in which each concept is contained. This method is adapted from Moed (2015), in that the simplest measure of a document's interdisciplinarity is the number of categories that a document cites, as seen in Figure 3. In this work, it is assumed that a document's title is a surrogate for the occurrence of concepts contained in the entirety of the document. Concepts that inhabit multiple conceptual spaces, in this case the spaces of NSF categories, are considered to be boundary crossing concepts. The number of categories in which each concept is contained is an indication of the measure of interdisciplinarity of a concept. These boundary crossing concepts were selected and analyzed for their frequencies of occurrence. For simplicity, concepts were first analyzed at the Grand Discipline to determine the overall layout of the core domain vs. the extended domain. The Grand Discipline is the least specific, or least granular, level of classes in the NSF taxonomy. Data were split in R according to their membership in either SSH or NSE, and by year. A script written in Python, located in Appendix A, was created to extract and quantify all concepts identified in each category. Concepts were then determined to be boundary crossing if they were present in both categories, which was determined using an inner join function in Pandas in Python.

To accomplish a boundary-contingent topical analysis, data were segmented both temporally and by high-level NSF category, resulting in twenty-two separate files. These files were each processed using the same Python script using the TextBlob library to remove stopwords and extract noun phrases and their frequencies of occurrence (Loria et al. 2018). Because "blobs" are essentially immutable types of lists, each blob was

Figure 4.1: Interdisciplinary Span of Concepts



exported from Python as a .txt file and re-imported as Pandas DataFrames containing 1) each noun phrase and 2) each noun phrase's frequency of occurrence. An inner join function was used to include only those noun phrases that occurred in both NSE and SSH, and output as a .csv including their frequencies of occurrence and respective categories. This file was imported and visualized in Tableau to create a "burst" analysis of topics that were of interest to both high-level categories, as well as their weights, over the eleven year time frame (Figure 5.3). For each disciplinary-level analysis, detected boundary concepts were extracted and aligned to their respective NSF disciplinary categories in a script in R.

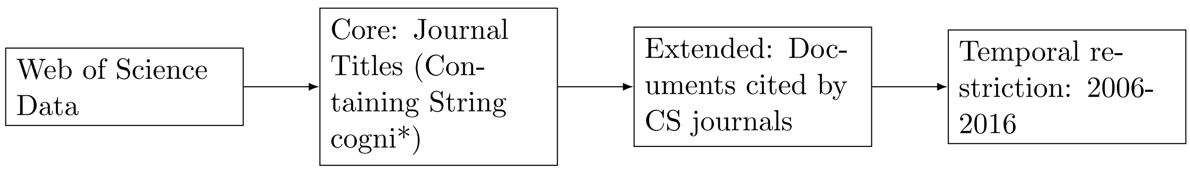
The topical structure of boundary objects was analyzed using only the core of the dataset, as analyzing growth for a seed and expand model would result in a decline in the overall frequencies of occurrence. The procedure and script used to analyze topicality on the entire domain of cognitive science were applied to the core set of titles.

4.4 Methodological Pipeline

The order in which the data was processed to follow the progression of research questions, or methodological pipeline, worked as a means on converging on highly interdisciplinary topics. The selection of data was critical, constraining the analysis to an eleven year period was done as outlined in Figure 4.2. The restrictions put in place to curate the final dataset were carefully chosen in part to make data management possible, and the

restriction of 2006 to 2016 applied to all works, both in the core and extended sets. This truncated the works cited of each core document to only those that were published in 2006 and after, which made the analysis of topicality, both overall and temporally, for Q1 possible in the way it was performed.

Figure 4.2: Data Selection



4.5 Focus Restrictions

This study does not address author-assigned keywords, journal-journal citation, pivot points, or issues of name disambiguation. Author-assigned keywords have been found to be inconsistent in their frequency of application across types of discipline, especially in humanities as opposed to STEM disciplines (Maurer and Shakeri 2016). It is unclear whether or not author-assigned keywords affect the assignment of indexer-assigned terms (Gil-Leiva and Alonso-Arroyo 2007). One study of documents contained in four separate databases found that 46% of all keywords assigned to a document additionally appeared with (21%) or without (25%) normalization as title or abstract descriptors (Gil-Leiva and Alonso-Arroyo 2007). Smiraglia (2015) found that the most frequently used terms in an article either align closely with indexer-assigned keywords, or vary greatly from those terms assigned by professional indexers. Additionally, author-assigned keywords in the Web of Science vary in consistency; more Keywords Plus are available in the Web of Science database than Author Keywords (Zhang et al. 2016).

Journal-journal citation is intentionally omitted in this analysis, as the scope of journal-journal citation is more used to reveal how highly cited an article is within its own, as opposed to other, fields (Pierce 1999). The focus of this research was on concepts and the authors who work on identified concepts.

4.6 The Cognitive Science Domain

As this dataset was curated in a seed and expand approach, each level, that is core and extended, are discussed throughout the results section. Comparing the core to the extended reveals the point of diversion between the number of works being published in the set of journals selected for the core, and the number of works pointed to retrospectively through the year of publication of the work cited. This was accomplished through subsetting data according to various axes (core vs. extended, SSH vs. NSE, and discipline-levels where appropriate for illuminating the structure of science in these areas) in R, the creation of tables using subsetting R's reshape2 library, and visualizing the resulting tabular data in Tableau.

4.7 Examining Boundary Concepts: The Domain of Cognitive Science

Mapping terms to the sub-domains in which they occur revealed a picture of the interdisciplinary conceptual content of a broad research area. From the entire domain, terms of high frequency that occur in multiple NSF categories were identified and traced temporally. Limitations of this method include that only terms contained in documents in the dataset can be traced, which excludes documents published in journals and by publishers not indexed by the Web of Science. Tracing of high-frequency concepts was done using Tableau to create a "burst analysis" of conceptual content as it appeared in each of the two high-level categories (NSE and SSH) over time. Timelines successfully convey multiple areas of examination chronologically in one chart, as well as frequencies of term use. Timelines can also be modified to show bursts of increased activity over time (Kleinberg 2003, Madsen et al. 2005). The "burst analysis" timeline created in this step contains the frequencies of occurrence of concepts that occur in both categories and in most or all years present in the domain (2006-2016).

4.8 Examining Boundary Concepts: The Cognitive Science Core

In order to examine the core, that is, journals the titles of which contain the string "cogni*," the domain was subsetted to contain only the core, then each grand discipline (NSE vs. SSH, respectively), and then again by year (2006-2016). This created another twenty-two files, each of which were processed using the script outlined in Appendix A. Output files were recombined into a single .csv file for analysis. To examine the growth in scope of the topicality of the domain, a time series analysis was conducted using R's built-in ts function for analyzing time-series objects. In this case, the unit of time was the year, and the unit measured was concepts in the core domain of cognitive science. R's HoltWinters() smoothing was applied to calculate the observed vs. fitted data to the smoothing model, and variance was calculated using the sum of squares.

Because this is a domain analysis focused on converging on highly interdisciplinary topics, other high-frequency boundary spanning topics, or those boundary concepts with the greatest interdisciplinary reach, were examined to establish aspects of aboutness of the domain other than interdisciplinary concentration, thus the first part of Question 2, "What boundary concepts are of the highest frequency?" This is examined for the entirety of the domain, as it gives a bigger picture of the domain's aboutness. To determine those concepts of the highest frequency, the count() function in R's dplyr library was used. Visualization of topics and temporal alignment was done in Tableau. The most logical resulting visualization of the temporal occurrence of boundary-crossing concepts in Tableau was a table.

Following Moed's simple measure of interdisciplinarity, the interdisciplinarity of a concept is the count of how many categories the concept spans, divided by the total number of categories under consideration. That is:

$$\frac{Numberofcategoriesaconceptspans}{totalconceptsindata} == 100\% \tag{1}$$

The topical periphera of 100% interdisciplinary boundary concepts provided insight into how the concepts were used in the context of titles. To align topics in categories, a

script was written in R comparing the list of boundary crossing concepts to the data.frame of core concepts, which was joined using rbind() to create a new data.frame with the 100% interdisciplinary concepts and their frequencies of occurrence per core category. The visualization of the topical periphera was done through a co-occurrence analysis of topics in titles, removing the node that represented the 100% boundary crossing concept, and then visualizing each network in Gephi.

Only concepts that were identified as 100% interdisciplinary from the core level of the domain were analyzed for the analysis of authors, their interdisciplinarity, and collaborations as exhibited through co-authorship, which is described in the next section.

4.9 Analysis of Authors, Interdisciplinarity, and Collaboration

As the motivation of this work was to examine the use of boundary concepts, identified boundary concepts served as the basis for the examination of author's interdisciplinary publishing habits. The number of authors publishing per year, and the number of authors contained in each category by year served as the basis for computations of descriptive statistics for the overall domain interdisciplinarity. Time series analysis was also implemented to examine the growth of authorship in the overall discipline, as well as the core vs. extended. The interdisciplinary spread of an author was calculated as the number of NSF categories in which each author published.

Author disambiguation has been a long-standing problem in bibliographic databases. Milojević (2009) noted the difficulty of identifying the precise number of authors in a database, instead, striving to mention the number of names contained in a database. Her solution was to treat authors with the same last names and different combinations of first initials as unique entities for the sake of analysis. Cases exist where authors do publish under different first names, such as "Tom" for "Thomas". Disambiguating names as much as possible allows for a clearer examination of patterns of collaboration, giving a clearer picture as to the underlying social structures that are present within and across disciplinary boundaries. However, authors whose names have changed may be missed, as tracing name changes without a master ID, such as ORCID, is nearly

impossible. Unfortunately, efforts to establish a disambiguation protocol are not united, making their coverage incomplete. These cases were remedied through pre-processing provided in the dataset, which contained not only the full author names, but each author's name standardized into the format "Last-FM." This format was selected for ease of use. Limitations of this approach include that common names, such as "Smith-J" all appear as the same author. However, the use of a concept-centric approach for the analysis of authorship restricts the scope of author data to those publishing on a given concept.

Frequency counts of authors by year provided the total number of publishing authors in both core and extended sets. Authors publishing in multiple high- and second-level NSF research areas were considered to be engaging in acts of boundary crossing. This was determined by calculating the number of times authors published, and whether or not they published in multiple NSF categories. Moed's (2015) measure of interdisciplinarity was applied as follows:

The percentage of boundary crossing authors was calculated as follows:

$$\frac{boundarycrossingauthors}{totalpublishingauthors} = percentage \tag{2}$$

The co-authorship analysis required performing extensive data manipulation in order to re-order the dataset in a way in which co-authorship analysis tools would recognize and process the data. Clusters of coauthorship were examined in different available tools (VosViewer and Bibliometrix). Because of the universal nature of the bibtex format, it was selected for reprocessing only the data needed into a new format. This required writing a simple script to process a .csv into a bibtex file and assigning dummy variables to preserve the NSF classificatory structure Ridenour (2020). A similar script was created to reorganize .csv data to the WoS ISI format. Both scripts, simple for loops in R, worked by assigning each column header to a variable, aligning it to the target output using the appropriate string formatting, then iterating over each row in the data.frame to achieve the target output. The dummy variables selected for preserving the structure of the NSF data were Acmid, Address, and Abstract, as these were not relevant for the analysis and enriched the data for .bibtex. For the ISI format, structural variables including NSF

categories were added back to the data after its analysis in Gephi.

VosViewer's output .gml file was imported into Gephi for co-authorship analysis, as the number of layout algorithms available to the program allowed for the selection of an algorithm that worked best to display the intellectual structure of each concept-centric co-authorship network. Authors who engaged in acts of boundary crossing through the act of publication in multiple categories as related to the concept-centric view were selected for further discussion.

Each largest connected component, as detected by VosViewer, was visualized and nodes were colored according to their cluster membership. This analysis was then contrasted with the a priori view of publication provided by the taxonomy when authors met the requirement of crossing boundaries through the act of publication on the key concept of analysis. VosViewer was selected instead of Gephi because the algorithms available maintained strict assignment of authors to detected clusters, instead of allowing for multiple membership.

Part of the underlying sociology of the domain was the number of authors per publication. This is calculated in a network as the average degree, which has a minimum of zero edges (indicate a sole-authored paper), and a maximum of edges (indicating the maximum number of authors in a network, which is $n + 1$.

4.10 Distinguishing Between Disciplines

Distinguishing between disciplines was done by divisions as determined by NSF classification that has been added at the journal level for the entirety of the database. This allowed for the imposition of a strict hierarchy on data, and for the ability to traverse downward a taxonomy to more specific assignments for a more granular analysis when necessary. The imposition of a strict classification system is necessary for empirical data-driven examination of boundary objects within information systems (Ridenour 2016). It is possible examination of co-occurring disciplinary assignment as evidenced through multiple assignment of WoS categories may provide a richer understanding of the intellectual structure of sub-domains within cognitive science.

4.11 Limitations

Strengths of this design include that it begins with a stable data source and a well-established classification scheme to use as a framework over which to identify areas of conceptual overlap as evidenced by terminology and citation. The scope of over 10,000 documents limits the specificity of the study to the broader transdisciplinary domain of cognitive science. Limitations of the analysis of the disciplinary span of boundary concepts include that a bag-of-words approach will be taken instead of determining each concept's proximity to cited sources based on the distance from the concept n-grams to citations as located in the text. A variation on this method was implemented in author co-citation analysis, and shows promise as a means to weight the interrelatedness of authors based on co-citation similarity in text Hsaio and Chen (2017). As this is a preliminary study that examined a large amount of textual data, I plan to address this in future research.

Additional limitations to the analysis was the way in which data was filtered to narrow in on concepts and authors. By focusing on the most interdisciplinary concepts at the disciplinary level (the second level of the NSF taxonomy), instead of the most frequently occurring concepts overall, the results show an interesting cross-section of what was most interdisciplinary for this particular domain. The methodology pipeline chosen to converge on the most interdisciplinary concepts in the core of the transdisciplinary domain limit the results, as each step filtered data to a more specific plane for analysis. Choosing to start at another point, for instance, just one set of selected journals, would result in an entirely different set of results.

The nature of the methodological pipeline converging on highly interdisciplinary core domain concepts, their conceptual periphera (as exhibited through term co-occurrence), and collaboration resulting in publication itself limited the possible results. Choosing to process data differently, examining concepts for properties other than being highly interdisciplinary in the core, would have resulted in a selection of different boundary concepts. Had concepts such as Dementia or Bipolar Disorder, which in of themselves are both *Diagnostic and Statistical Manual of Mental Disorders* and *International Classifica-*

tion of Diseases billable and coded disorders, the inhabitance of the concepts and their alignment to the NSF taxonomy would be very different than the two concepts chosen based on their interdisciplinary nature. Choosing to analyze the data from a perspective of methodology, which would literally be how researchers know, would have resulted in a periphera of conditions and populations to which they were applied.

Analyzing collaboration is in of itself tricky, as not all work done is published. Much work that is done, including the serendipitous moment of inspiration when an individual makes hidden connections to develop understanding, is either not documented, or simply intangible. Collaboration, either within one's home discipline or organization, or across institutional boundaries, does not necessarily result in publication. For the purposes of this analysis using bibliometric data, only published works within the Web of Science for the particular subset in the transdisciplinary domain were selected. Co-authorship trends, such as the number of works published and the number of authors per paper, are concrete. The underlying sociology, evidenced through published co-authored works, would be different depending on the concept chosen for analysis.

4.12 Selection of Methods

Unlike social research mixed methods, in which a multi-trait multi-method matrix for calculating convergence on a single phenomenon from multiple measures is appropriate Campbell and Fiske (1959), each method used in this study to converge on conceptual overlap (natural language processing, co-occurrence analysis, co-authorship) was chosen to analyze a separate phenomenon to converge on facets of the same phenomenon (terminological use, sources, and authorship patterns). Convergence on boundary concepts was determined by concepts residing within a document by the examination of the document's title and the co-authorship patterns. This step-by-step approach allows for a refinement of analysis through the examination of topical, temporal, and authorship changes in a big data dataset.

4.13 Conclusion

The complex, and polycontextual, sociocultural navigation required for successfully engaging in inter- and transdisciplinary research and publication can be identified in a citation database by examining the following against a classificatory framework: the natural language used to describe concepts in titles and abstracts; citations made to works published in other fields; and patterns of authorship, both co-authoring and acts of boundary crossing by publishing in multiple disciplinary categories. The methodology I have outlined examines encoded language, topical co-occurrences, and authorship; it leverages the available structures to identify concepts, corresponding key documents, and their authors whose publications reside in shared intellectual space. This approach does not neatly fit the research traditions of knowledge organization, information retrieval, or informetrics; instead, it uses core principles and methods from each to provide a clear understanding of a transdisciplinary oeuvre of research.

The steps selected, outlined in this chapter, work to converge on highly interdisciplinary concepts (those that span all six core boundaries, and are thus discussed as both boundary concepts, and boundary-spanning concepts), the conceptual periphera (as shown through term co-occurrences with highly boundary spanning concepts), and those published authors who worked on identified highly interdisciplinary concepts. Multiple other directions were possible, which are discussed in Chapter 7: Future Work. Each piece of this methodology was crafted to analyze the inner workings of cognitive science, explicitly showing topicality as opposed to a further abstraction of bibliometric data.

Chapter 5

Results

5 Results

Boundary objects can be identified through the use of data science methods. The transdisciplinary domain grew over the eleven year period of 2006-2016, and saw an increase in the number of authors publishing in the core of the dataset (those journals which contained the string "cogni*"). The greater dataset, including the core and extended, decreased in the number of publications over time. This is to be expected, as work can only be cited which has been done and published in the past, and the selection of publications was based on journals containing the string cogni* and those works that the papers cite.

5.1 Epistemological Boundaries

1) What epistemological boundaries, manifested as classificatory divisions, exist between NSF-identified disciplines that comprise cognitive science? What is the representation of work published in these? How does this change over time?

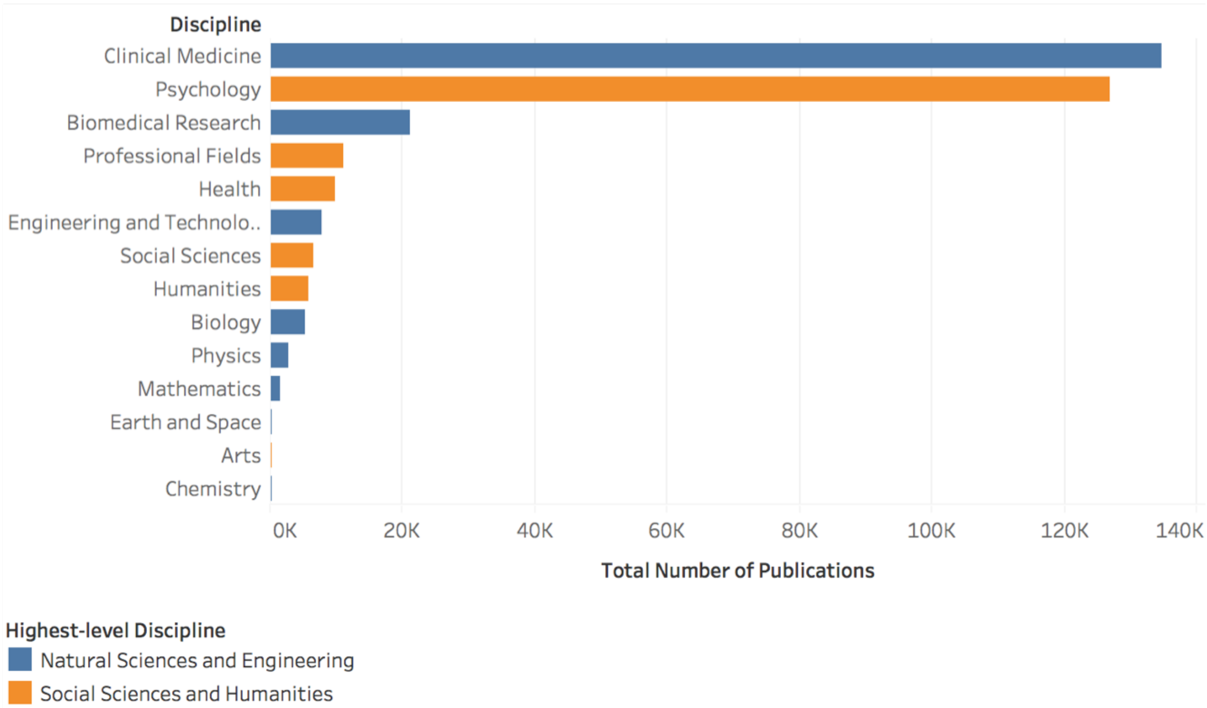
To answer the first part of this question, the epistemological boundaries used are artificial, as they are manifested by a classification. I argue that the taxonomy is a representation of identified epistemologies within like-journals, as each node in a taxonomy contains carefully identified self-similar ways of knowing about common research interests. These include methods, which are in of themselves ways of knowing. This logically follows, as disciplines are social, and methodology is a socialized way of knowing. At

the first level of the NSF taxonomy (coded as GrDisc), Social Sciences and Humanities (SSH) and Natural Sciences and Engineering (NSE) are the highest-level epistemological boundaries between subsets of the domain. At the second "disciplinary" level, a more specific nonmenclature was used to describe the cognitive essence of the work contained under the taxonomic penumbras of SSH and NSE. In the core, SSH consisted of Arts, Social Sciences and Humanities consists of Humanities, Psychology, and Professional Fields. Core NSE disciplinary boundaries were drawn around Biology, Clinical Medicine, and Engineering and Technology. The extended set included six disciplinary divisions in SSH: Arts, Health, Humanities, Professional Fields, Psychology, and Social Sciences; NSE contained eight disciplinary divisions: Biology, Biomedical Research, Chemistry, Clinical Medicine, Earth and Space, Engineering and Technology, Mathematics, and Physics.

The second part of question 1, regarding the representation of work published in these divisions, follows. In the case of the entire domain of cognitive science, the meso-level of publications inhabit SSH. In order to determine the frequencies with which publications aligned along the conceptual framework of the NSF taxonomy, the second level of the taxonomy was used as the framework against which to calculate the concentration of publications. The number of articles published per NSF taxonomic category can be seen in Figure 5.1 and Table 5.1.

The core of the domain consists of six NSF disciplines that are themselves subcategories for two grand disciplines. This core contains a total of 29,006 articles: 1) Biology (991 articles), Clinical Medicine (8,286 articles), and Earth and Space (446 articles); and 2) Humanities (786 articles), Professional Fields (147 articles), and Psychology (18,350 articles). These six disciplinary spaces contain all publications with articles the title of which contain the string "cogni*," constituting the main epistemological spaces for cognitive science works contained in the WoS database. As disciplines are closely related to domains, and as the terminology is sometimes used interchangeably, each of these six areas was treated as a domain unto itself. Thus, all journals of which the core of the transdisciplinary domain consists are contained in these disciplines. The distribution of publications in these disciplines is heavily concentrated in the SSH discipline of Psy-

Figure 5.1: Epistemological Spread of Cognitive Science: Discipline Level



chology (63.263%), and the NSE discipline of Clinical Medicine (28.567%, see Table 5.1). Thus, most work done on cognitive science topics inhabits the epistemological information space of the database in Psychology and in Clinical Medicine.

To answer the third part of question 1, Table 5.1, breaks down the number of publications by the axes of 1) core (C), extended (E); 2) Grand Disciplines (Gr.Disc) of Natural Sciences and Engineering (NSE) and Social Sciences and Humanities (SSH); by discipline (Disc), and by year. Examining publications per category by year revealed overall trends in publications (Table 5.2). The most works in the cognitive science domain were published in Psychology and Biomedical Research. The number of publications in each of the fourteen second-level categories showed increases and decreases in publication output in variation that would match Bradford's law, as not all areas will create the same number of documents in a given span of time.

Specifically, the number of core publications increases yearly. In NSE, a total of 2,019 articles were published in 2006, and 2,690 were published in the core of SSH. By 2016,

Table 5.1: Num. Publications by Level, by Gr. Disc, by Disc, by Year

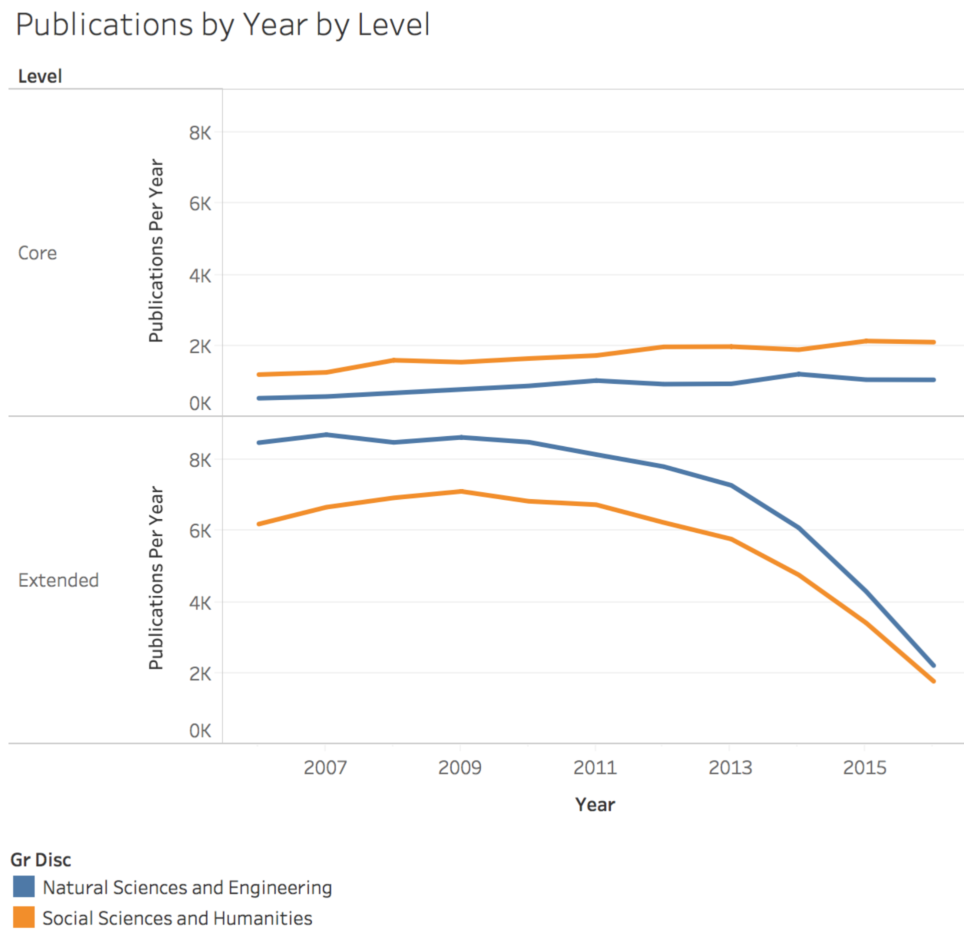
			Year										
L	Gr Disc	Disc	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
C	NSE	Bio	36	47	73	88	80	87	116	92	131	127	114
		CM	472	514	591	672	766	930	776	767	1040	881	877
		E&S	23	20	22	23	29	31	49	78	56	50	65
	SSH	Hu.	15	25	52	81	84	88	81	106	86	90	78
		Pr.Fi.	13	12	12	14	13	15	13	17	12	11	15
		Ps.	1177	1234	1556	1464	1561	1675	1909	1872	1812	2074	2016
E	NSE	Bio	255	248	235	217	199	200	162	150	109	79	24
		B.Res	805	1019	1034	1204	1046	1251	1384	1198	1001	590	220
		Ch.	9	10	8	9	11	6	12	11	5	5	1
		CM	6590	6632	6412	6306	6256	5816	5244	4652	3586	2144	734
		E&S	12	20	13	13	14	22	20	13	25	14	6
		E&T	419	340	385	343	386	351	311	317	329	206	80
		Ma.	64	84	102	110	106	91	90	59	65	33	11
		Ph	171	163	165	148	100	105	97	69	31	35	6
	SSH	Ar.	10	8	14	12	4	11	8	4	5		2
		He.	457	506	514	534	519	477	414	397	218	163	42
		Hu.	225	263	304	280	281	303	212	207	139	85	27
		Ph.	504	461	510	529	494	442	403	342	241	157	54
		Pr.Fi	4571	4945	5097	5186	4967	4987	4519	4004	3208	2013	711
		S.Sc	271	332	337	302	267	211	246	204	157	95	30

Table 5.2: Number of Core Publications Per Year by Category

Year											
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
NSE	9011	9284	9167	9409	9373	9174	8733	8217	7295	5345	3257
Bio	300	309	309	313	288	289	284	257	258	228	151
B.Res	812	1039	1045	1235	1094	1302	1450	1354	1160	790	437
Ch.	9	12	8	10	12	5	13	12	5	5	1
CM	7170	7260	7090	7165	7295	6924	6330	5961	5250	3847	2366
E&S	13	21	14	13	16	22	24	18	25	19	16
E&T	466	388	424	400	450	426	419	461	473	364	239
Ma	66	89	108	114	107	96	100	68	79	44	21
Ph.	175	166	169	159	111	110	113	86	45	48	26
SSH	7387	7922	8528	8655	8480	8468	8210	7751	6656	5547	3875
Ar.	10	10	14	13	7	12	12	6	9	1	3
He	472	525	535	551	557	499	450	441	287	210	81
Hu.	240	300	367	378	384	420	331	353	259	215	148
Pr.Fi	545	499	547	590	555	513	469	452	322	250	141
Ps.	5834	6245	6707	6781	6671	6788	6664	6263	5575	4731	3434
S.Sc.	286	343	358	342	306	236	284	236	204	140	68
	16398	17206	17695	18064	17853	17642	16943	15968	13951	10892	7132

this increased to 4,297 in NSE and 5,889 in SSH. Despite more articles being published in SSH, more articles were cited in NSE, as can be seen in Figure 5.2.

Figure 5.2: Publications by Level by GrDisc by Year



5.2 Terminology and Boundaries

1A) How do terminological boundary objects fit against the NSF's classificatory framework? What concepts are the most interdisciplinary in the core of the domain?

To answer the first part of Question 1A, terminological boundary concepts identified were those that inhabited both NSE and SSH. The entire domain, that is, those articles published in the entirety of the seed and expand sample, contained a higher concentration of articles published in NSE than in SSH. This frequency is skewed in part due to the number of second-level classes present in the areas in which CS publications exist within journals in the taxonomy; NSE contains eight subclasses, while SSH contains six

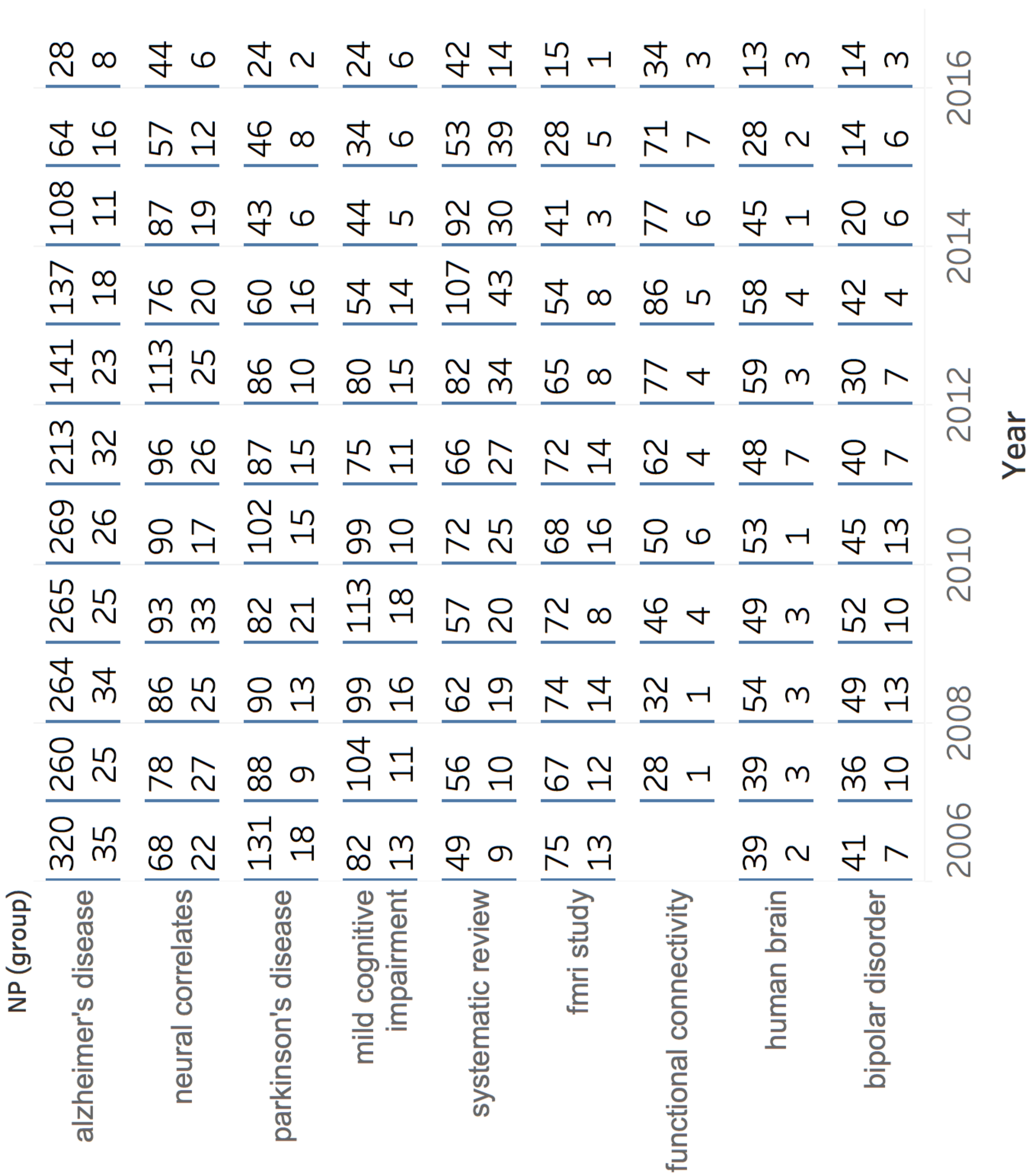
subclasses. In the core of the dataset, this distribution is three and three.

A total of 7,149 noun phrases were determined to be boundary concepts in the domain of cognitive science. The initial scope of the data, which included 202,132 records over 25 years, was reduced to an eleven year span including 176,034 records. This reduction was a conscious choice in order to engage in a more in-depth analysis of boundary crossing of concepts. Additionally, this reduction allowed for the examination of the crossing point between citing works (the core, or seed of the domain) and cited works (the expand). This cross-section was identified as 2012, as the total number of publications in the core continued to increase while the number of works published in that year decreased.

Terminological boundary concepts that fit the criteria of publication within both NSE and SSH for ten or more years included Alzheimer's disease, neural correlates, Parkinson's Disease, mild cognitive impairment, fmri study, functional connectivity, human brain, and bipolar disorder (see Figure 5.3). These terminological boundary objects were not necessarily the most frequently occurring overall, but instead were the concepts which spanned both high-level taxonomic boundaries. The distribution of concepts is not even, with a greater frequency of each boundary concept in NSE. This naturally follows from the uneven distribution of the transdisciplinary domain, which contained more publications in NSE than in SSH, which is to be expected given the distribution of publications within the taxonomy. The visualization in Figure 5.3 is akin to a burst analysis in Sci2 software, with three important distinctions: 1) only boundary inhabiting concepts were included in the analysis, 2) the frequencies of concept occurrence in titles are accounted for numerically, and 3) the creator of the visualization, not an algorithm, determined the appropriate cutoff threshold for visualization. In the case of Figure 5.3, only those concepts that occurred more than 53 times and occurring for at least nine out of the eleven years studied were included. In Figure 5.3, the top number is the count of boundary concepts occurring in NSE, and the number below is the count of boundary concepts contained in SSH .

To answer the second part of Question 1A, the most interdisciplinary concepts in the core of the domain are "children's," "case study" at 100% interdisciplinarity. "Cognitive

Figure 5.3: Boundary Concepts: Burst Analysis



control," "individual differences," "social cognition," "visual representation," and "young children" were boundary-spanning concepts present in five out of six core disciplines, at 83.33% interdisciplinarity.

5.3 Core-level Interdisciplinary Concepts

1B) "What boundary crossing concepts are of the highest frequency? What boundary concepts at the core-level are 100% interdisciplinary, and what are their topical periphera? How do boundary-crossing concepts in the core change over time?"

There was a steady increase in boundary-spanning conceptual growth over time (see both Figure 5.4 and Table 5.3). As the growth pattern demonstrated an upward trend, an additional analysis was performed to calculate the predicted growth in the frequencies of occurrences of concepts (Figure 5.5). This is not the best indication of actual trends as the number of journals publishing in the core changed over time; however, it demonstrates a growth in the acceptance of new topics in the core domain of cognitive science.

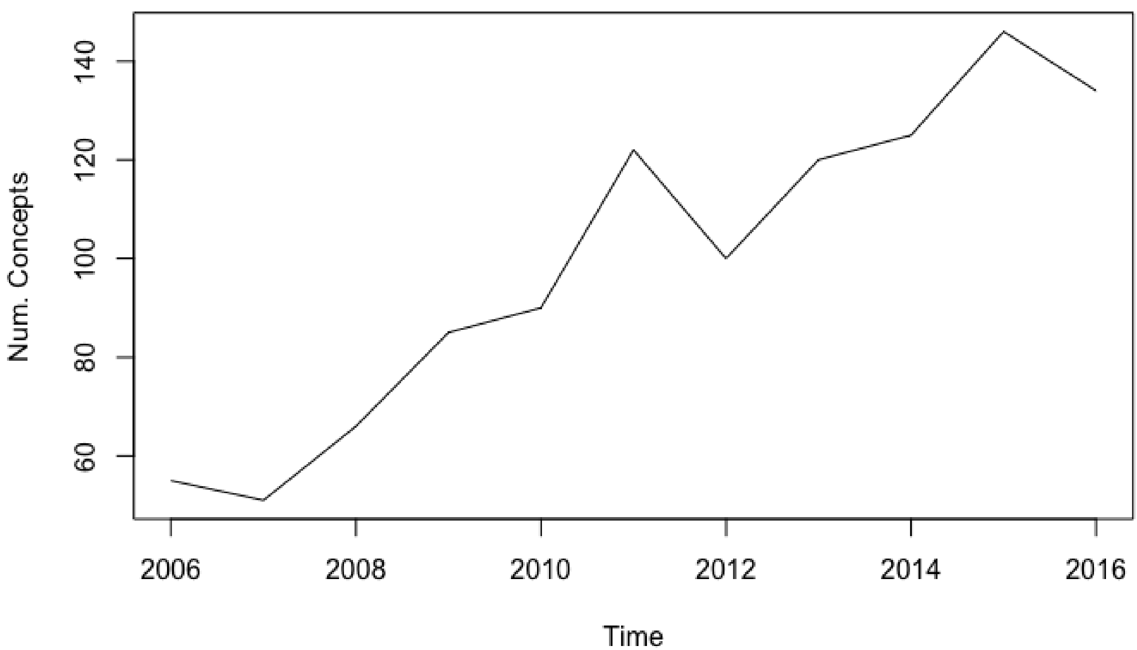
Table 5.3: Number of Boundary Concepts per Year: Core Data

Year	Total Concepts
2006	55
2007	51
2008	66
2009	85
2010	90
2011	122
2012	100
2013	120
2014	125
2015	146
2016	134

To answer the temporal aspect of Question 2, the number of boundary concepts per year was calculated and visualized as a table (Table 5.4). Only the boundary-spanning concepts occurring 16 times or more were included in Table 5.4. The number of boundary-spanning concepts was visualized as a graph (Figure 5.4), and a time-series analysis of the growth of the number of boundary concepts for the domain was performed. The overall

frequencies (core and extended) are discussed in this analysis to provide context to that which occurs in the core.

Figure 5.4: Total Number of Core Boundary Concepts per Year



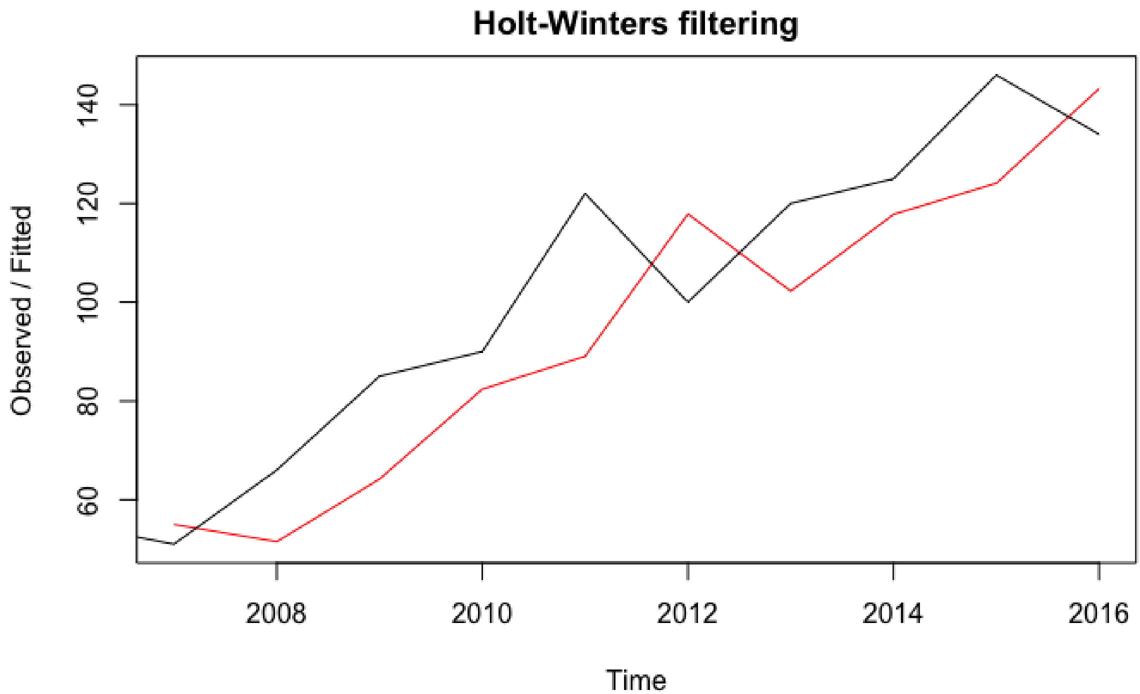
As is visible in Figure 5.4, steady growth in the number of concepts present as noun phrases in titles of the core of the domain is visible. Additional analysis to measure the predicted growth in the number of concepts demonstrated that the predicted trend for concepts was continued growth through 2016, instead of the slight decline between 2015 and 2016, from 146 concepts to 134 concepts. The predicted curve was calculated for only the time period of the actual data using R's HoltWinters() function, and is shown in Figure 5.5, where the black line is the actual observations, and the red shows the predicted trends. Exponential smoothing for the model used an alpha of 0.875. Alpha is a value between 0 and 1. Zero is indicative of less weight being placed on recent observations for the forecasting of future values, whereas 1 places more weight on recent observations. The weight of 0.875 demonstrates that more recent frequency values were more meaningful in predicting future trends for conceptual growth using this measure. The measure of the model's accuracy was calculated by the sum-of-squared errors, which

Table 5.4: Number of Core Boundary Concepts per Year Occurring 16 or More Times

Concept	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
individual differences	10	17	11	14	16	28	22	22	35	26	31
alzheimer's disease	32			40	29		22	14		15	9
mild cognitive impairment	20	8			18	8	12		11	7	11
neural correlates		7	9	12	11	13	14	13		15	
cognitive control	2		7	7	6	12	4	13	14	11	9
social anxiety				11				12	10	16	14
depressive symptoms		6	4			4	10	9	10	10	9
decision making	3	2	6	4	6	4	5	6	4	11	8
time course	3	4	3	8		6	10	7	5	5	8
recognition memory	8		6	10	5	9		5	6		7
emotion regulation	2	4		3		9	5	9		9	12
social cognition			4	5	4		7	6	8	5	9
case study					5		6	9	7	5	11
cognitive performance	4	4	3		5	7	3		5	5	6
memory capacity				4	8		9	7	2	12	
neural basis	3	2		5	5	7	6		8	4	
prospective memory	4						10	9	9		8
young children				5		5	8		4	5	13
eye movements	6	4	14			6	8				
episodic memory	2			8	6	4	4		3	5	5
computational model		6	4	3	3	6		5	5		4
erp study		3	3	6		5		6		6	6
pilot study	3		5	6			4	6	5		5
functional connectivity				2		4				14	13
systematic review				3	2		2	5	6	4	11
visual search				2	8	9	3			7	4
mental rotation			4	4	5	5		9			3
selective attention			3		2	6	7	4		4	4
executive function	3		2				5	8		4	5
facial expressions	2		7	2			5	4		2	4
memory performance					5	5	3	3	2	4	4
word recognition	2		3	4		2	7		4		3
parkinson's disease				4		4	6	7		4	
attentional control	2	2			2	3	3		4	2	6
young adults				3		2	6		6	7	
visual attention	4				4	4	2			7	
executive control	4				3	3		7	3		
false memories					6	8			6		
functional magnetic resonance		12		3				2		3	
emotional stimuli			2			4	4		5	4	
executive functions							5		5	2	7
macaca mulatta		3				3		2	4	5	2
pan troglodytes				15				4			
short-term memory		3			5	3	4		2		2
attentional bias					4	2		5	7		
cognitive function			3						5	4	6
electrophysiological evidence				5		4			4	2	3
implicit learning				4	9						5
inhibitory control						4			9	2	2
obsessive-compulsive disorder	2	4		6					3	2	
spatial attention					3	7	3	4			
williams syndrome	4	2		4	2	5					
different types					4			2	5		5
language acquisition					4	2	4			6	

calculated as 3055.885. The high sum of squares value indicates there is a great deal of variation in the data.

Figure 5.5: Total Number of Core Boundary Concepts per Year: Fitted Time-Series Model



5.4 Results: Entire Domain

Publications peaked in 2009 with 18,064 articles, and steadily decreased until 2016 (Table 5.5). Although the number of overall publications decreased, this is an artifact of using a seed and expand method to gather citation data, as work can only be cited that has been done in the past. This is clearly seen in Table 5.6.

5.5 Core: Analysis of Concepts

To answer the first part of Question1B, "What is the broadest disciplinary spread of terminological concepts present in the dataset?" only two concepts. This analysis also establishes the basis for the analysis of question 2, "What are the underlying sociology (through acts of co-authorship)? What acts of author boundary crossing exist sur-

Table 5.5: Number of Publications per Year (Core and Extended)

Year	Num.
2006	16,398
2007	17,206
2008	17,965
2009	18,064
2010	17,853
2011	17,624
2012	16,943
2013	15,968
2014	13,951
2015	10,892
2016	7,123
total	172,044

Table 5.6: Article Count by Year by Level

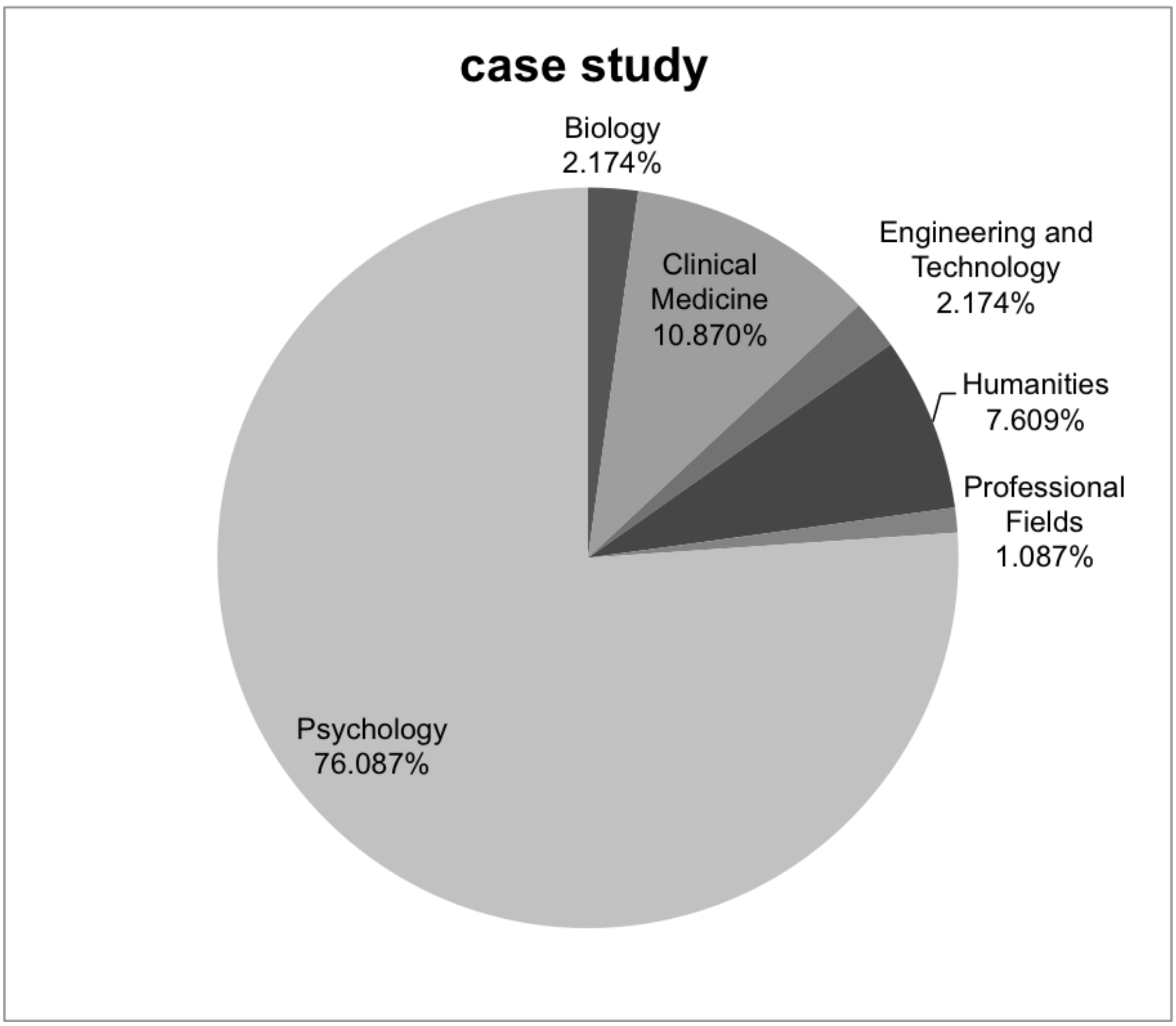
Year	Core	Extended
2006	1736	14821
2007	1851	15568
2008	2306	15684
2009	2341	15892
2010	2533	15447
2011	2826	15210
2012	2941	14298
2013	2926	13171
2014	3136	10993
2015	3234	7861
2016	3192	4030

rounding the most interdisciplinary concepts?" as it identifies the most interdisciplinary concepts in the core of the domain. A total of 657 concepts were computed in the core dataset; however, only 655 were located in core title data. Closer examination of the data as it was distributed at the disciplinary level of the taxonomy revealed that the majority of concepts occurred in Clinical Medicine (28.801%) and Psychology (64.621%). This is the same result as the number of concepts in the core and extended dataset. Cognitive Science topics in Clinical Medicine were largely about Alzheimer's Disease, cognitive impairments, neural correlates, and age differences. Psychology was largely focused on "new evidence," children, ERP evidence, and cognitive behavioral therapy.

Of these 657 concepts, two were present in all six disciplines. To examine the con-

ceptual periphera of highly interdisciplinary concepts, the co-occurrence of each concept with that of other concepts by document was made into individual .csv files per concept. This .csv file consisted of the record ID and the title data for every record containing the concept being examined, thus, one for "children's," and one for "case study." Co-occurring terms with a minimum threshold of two occurrences were selected, and default clustering was used to pre-process the data in VosViewer. The file was saved as a .gml file, which is a type of open network file structure that allows for more metadata to be associated with the network file than that of a .net file. The .gml file all the publication records containing each topic was then imported into Gephi, which allowed for greater control of the resulting visualization.

Figure 5.6: Core: Case Study Distribution of Disciplines



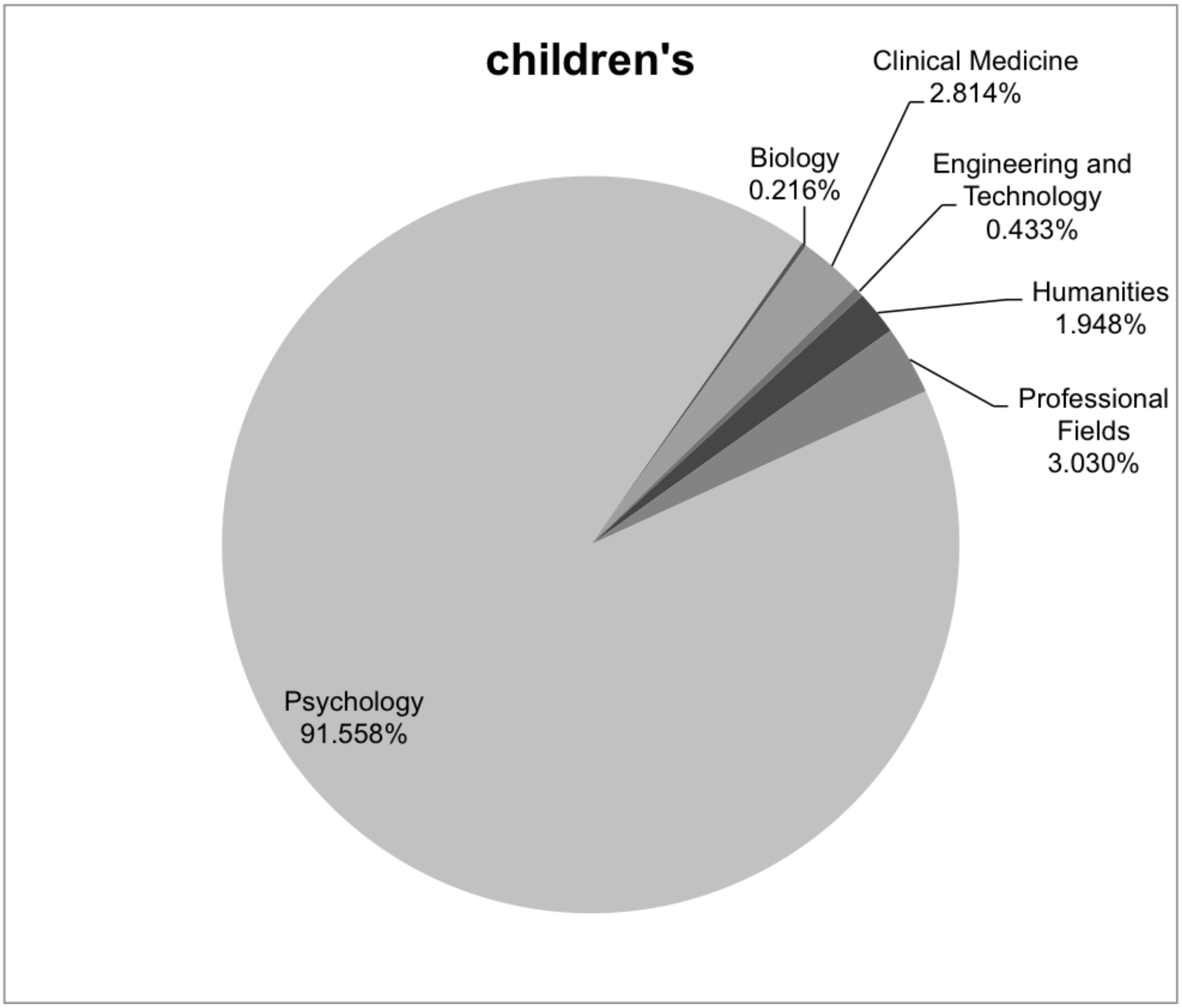
Concepts that occurred in all six disciplines in the core level of data in Cognitive Science included the following: case study (Figure 5.6), and children's (Figure 5.7). "Case study" occurred a total of 92 times. The majority of occurrences of "case study" were published in the core discipline of Psychology (70 times, or 76.086%), followed by Clinical Medicine (10, or 10.8%), Humanities (7), Biology (2), Engineering (2), and Professional Fields (1). The distribution of the frequencies of occurrence of the concept "case study" follows that of a long tail distribution, or power law. In this case, the number of boundary crossing concepts per category.

The concept "children's" occurred a total of 492 times in the core of the domain. The greatest frequency of occurrence of the concept "children's" was 423 times (91.558% of occurrences), in the category of Psychology. The next most frequent occurrence was 14 times in Professional Fields (3.03%), 13 times in Clinical Medicine (2.814%), 9 in Humanities (1.948%), twice in Engineering and Technical Fields (0.433%), and and once in Biology (0.216%). The greatest concentration of occurrences of the concept "children's" is concentrated in Psychology.

"Children's" for co-occurring concepts in the entirety of the domain resulted in an extremely large graph with 2,444 nodes and 637 clusters. The number of clusters followed the expected long-tail distribution, with the largest cluster containing forty-two concepts and the smallest containing only one. Co-occurrences detected in VosViewer were exported to Gephi for greater control. In the case of "children's," all nodes and edges were kept in the network in VosViewer in order to examine the highest frequency occurring clusters in Gephi. Had this been done in VosViewer, it wold not have been possible to examine the entirety of network components as clusters of higher-frequency co-occurring concepts because any sampling reduced the number of co-occurring terms. Co-occurring concepts with a threshold of occurring twice were included in the initial visualization, which was refined in order to display the conceptual periphera more clearly.

The network visualization for "children's" was created with nodes of a degree of 5 or greater, which included a total of 232 nodes, or 9.49% of nodes, and 360 edges, or 11.46%. Co-occurrence graphs are undirected, meaning that the edges of the graph convey more

Figure 5.7: Core: Children's Distribution of Disciplines

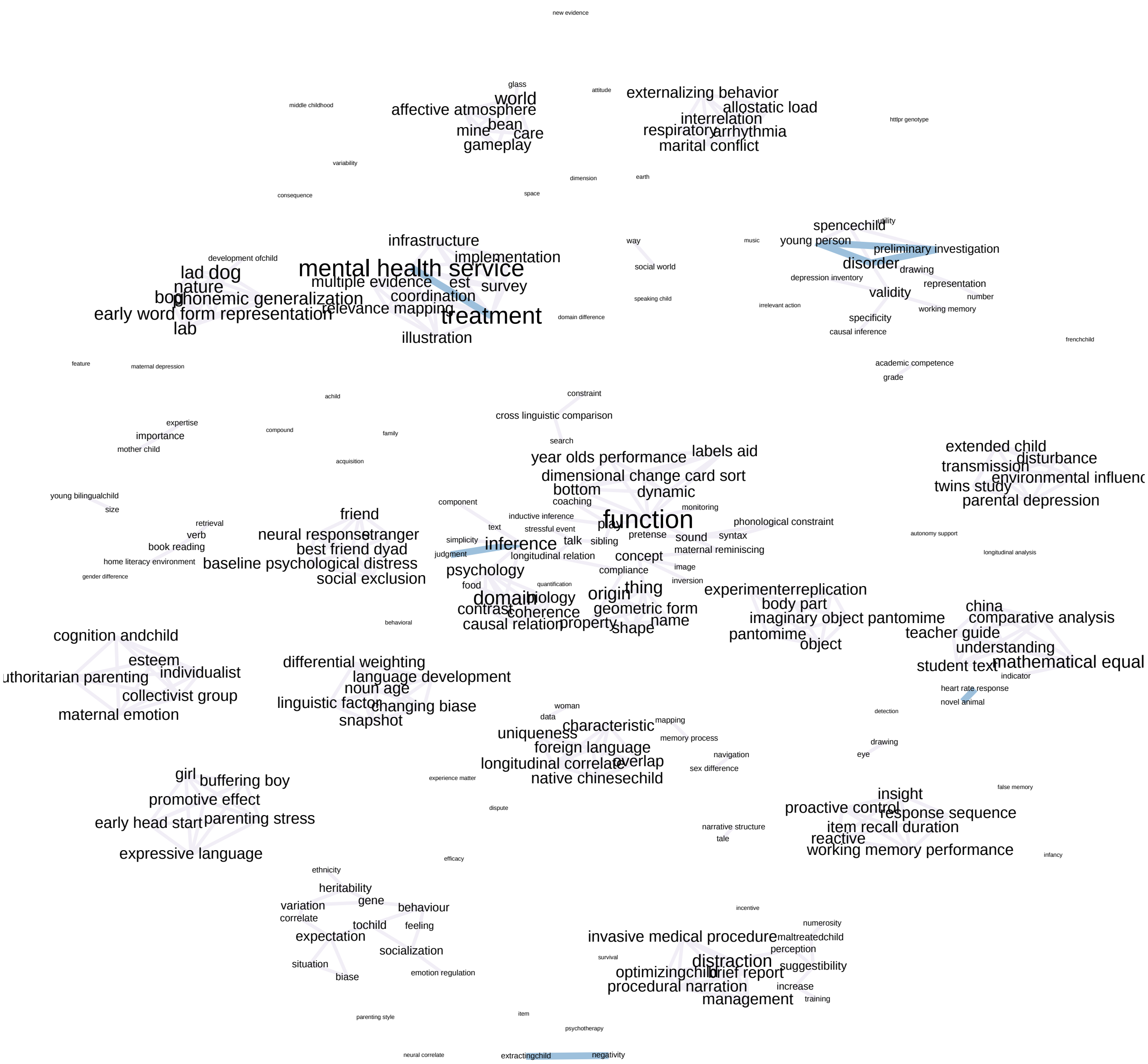


meaning than the total number of connections of each node. As such, the visualization was created using only edges and labels to demonstrate the connections between component concepts.

The network visualization of co-occurring concepts in titles for Children's clearly shows topical clusters including (but not limited to) functions, mental health services, gender, environmental influences, and understanding 5.8. In all, 637 clusters were identified. For the network visualization, the graph was filtered to limit the degree range to at least 5, meaning that the each node displayed had a minimum of five connections, or co-occurrences, to other nodes in the graph. The maximum degree in this co-occurrence network was 23.

Concepts that frequently occurred with the boundary-spanning concept "children's"

Figure 5.8: Children's: Co-occurring Concepts

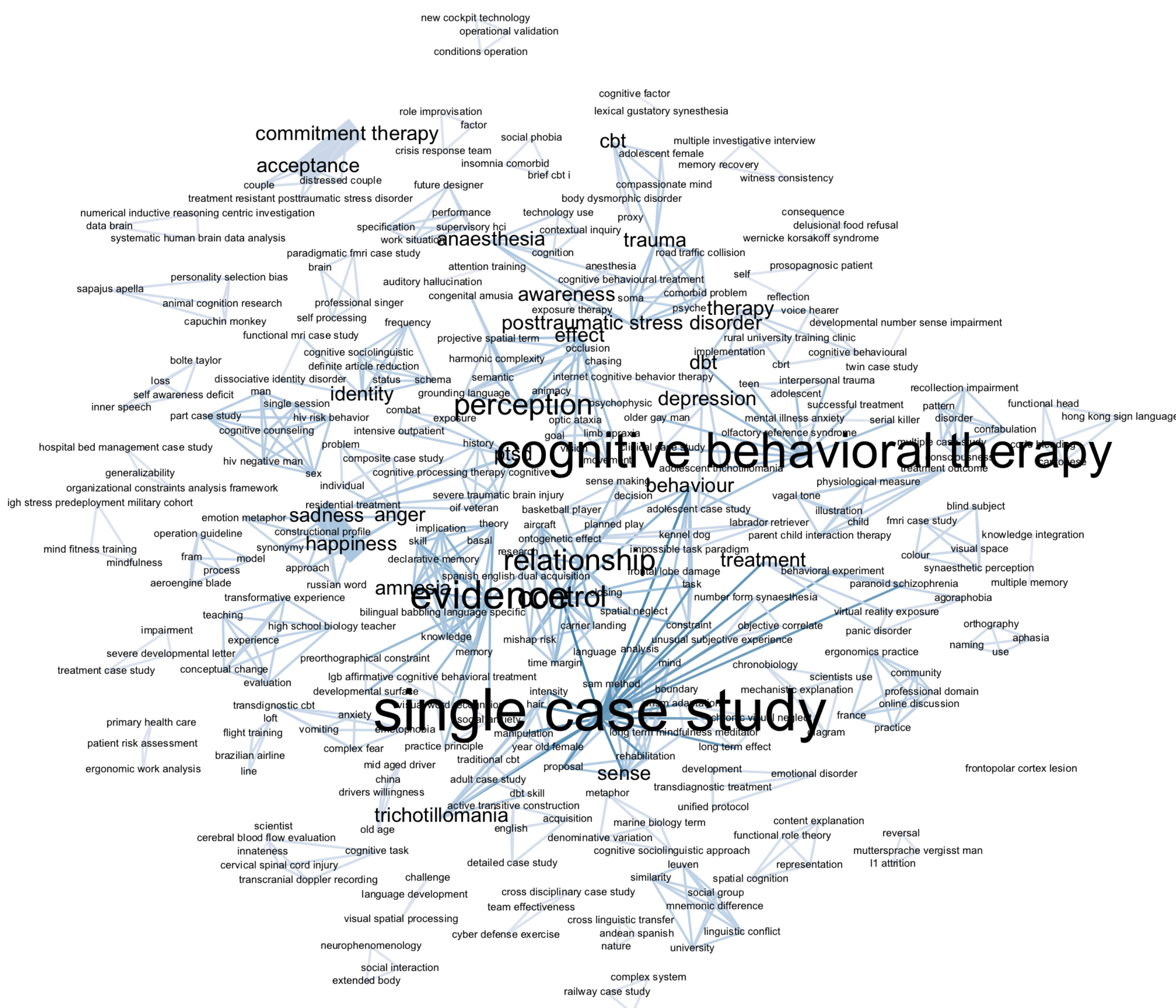


included function (25), judgment (18), inference (15), representation (14), concept (12), object (11), false memory (10), drawing (8), grade (8), situation (8), talk (8), development of child (7), disorder (7), and monitoring (7). Concepts occurring six or fewer times were not included for clarity.

In all, 3,479 authors contributed to"children's" concept publications. Of these, 740 authors published more than once. A total of 382 authors published in both the core and the extended sets, spanning the boundary of core to extended. Eighty-two authors published in both NSE and SSH.

The concept "case study" had 329 co-occurring concepts and 60 clusters (5.9). The graph contained 884 edges. The OpenOrd layout was selected for visualizing this graph because of the clarity with which it displayed the co-occurrences of concept terms with case study, and only edges occurring at least three times were included in the visualization. As is clearly visible in the visualization, "case study" is applied to several research problems and populations within the transdisciplinary domain of cognitive science. Linkages between connected components included concepts which constitute types of evidence (e.g. evidence), clusters of related concepts (such as emotions including sadness, happiness, and anger), and ways of managing conditions (e.g. treatment).

Figure 5.9: Case Study: Co-occurring Concepts



5.6 Authorship Analysis

2) What are the underlying sociology (through acts of co-authorship)? What acts of author boundary crossing exist surrounding the most interdisciplinary concepts?

Following the restrictions set in place for the analysis of data through a methodological pipeline, authorship was examined last. Overall trends for the core vs. the extended for each higher-level NSF discipline showed an increase in the number of publishing authors in the core of the discipline (Table 5.7). What appears to be a decrease in the number of publishing authors in the extended set and domain is the result of the implementation of a seed and expand method for data selection. This is the result of the lag in time between the act of publishing and the act of citation of publications.

5.6.1 Author Boundary Crossing

To establish context for the first part of Question 2, an outline of the overview of the publication frequencies in total, the average number of authors per article, and the number of authors publishing in each top-level NSF category. In total, 210,383 authors published a total of 160,405 articles. The mean number of authors per publication was 1.312. Minimum numbers of authors per article was 1, and the maximum was an outlier, attributed to 3,220 authors. This paper, Article ID 45964150 was published in the NSE discipline of Physics. Because of the high occurrence of coauthors, it was omitted for the co-authorship analysis. A total of 19.964% of authors engaged in acts of boundary crossing through the act of publishing in both NSH and SSH categories.

A total of 88,471 authors in the domain published more than one article. In total, 121,912 authors published only one article contained in the domain analyzed. In NSE, 163,117 authors published a total of 82,769 articles. The average number of authors per article per publication in NSE was 1.855. SSH contained 89,265 authors who published 77,636 articles. The average number of authors per publication was 1.250. The difference in the number of authors per publication follows what would be expected, as it is well known that more collaboration is done in "hard" sciences than in humanities disciplines.

The total number of publishing authors increased in both NSE and SSH (Table 5.7).

Table 5.7: Publishing Authors by Year by Category by Level

Cat.	Lev.	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
NSE	C	2019	2206	2676	3029	3323	3771	3571	3624	4708	4128	4297
	E	27558	29397	29402	29376	33274	29211	27109	24964	21169	13615	5008
SSH	C	2690	2978	3710	3678	4011	4337	4899	5043	5154	5750	5889
	E	14214	15913	16633	17316	16684	16624	15488	14419	11429	7677	2723

In NSE, the core increased from 2,019 authors to 4,297 authors; in SSH, the core number of publishing authors increased from 2,690 authors 2006 to 5,889 in 2016.

5.6.2 Co-authorship Surrounding Core Boundary Spanning Concepts

To answer the second part of question 2) regarding the sociology surrounding boundary concepts as evidenced by co-authorship for each of the most interdisciplinary core topics, a concept-centric analysis of co-authorship in the transdisciplinary area of cognitive science was applied to the subset of authors who published on the concepts of "children's" and those publishing on "case study." This provided an insight into the underlying sociology surrounding work done on these two highly interdisciplinary concepts. To examine the underlying sociology, social network statistics were calculated using Gephi, and are displayed in Table table:childrensnetworkstatistics. The average degree of the network, indicative of the social connections per paper is seen through the network's the average number of connected nodes. was 3.628, with a minimum of zero (single-authored publications), and a maximum of 21 (22 co-authors on a publication).

The network diameter was 15, indicating that the longest linkages of coauthorship (in terms of shortest paths) was 15 edges between the first author in the chain and the last author in the chain. Children's coauthorship graph density was 0.003, which is a ratio between the number of edges present in the network to the possible number of edges (Wasserman & Faust, 1994). The modularity, or clustereredness, of children's is .979.

The children's co-authorship network contained 1,101 authors. These can be seen in their entirety in Figure 5.10, displayed using the Fruchterman-Reingold layout in Gephi. This provided a macroscopic view of the frequency with which authors, no matter their area, publish on an interdisciplinary topic. Interestingly enough, the authors with the greatest number of publications in the selected time frame were not necessarily those

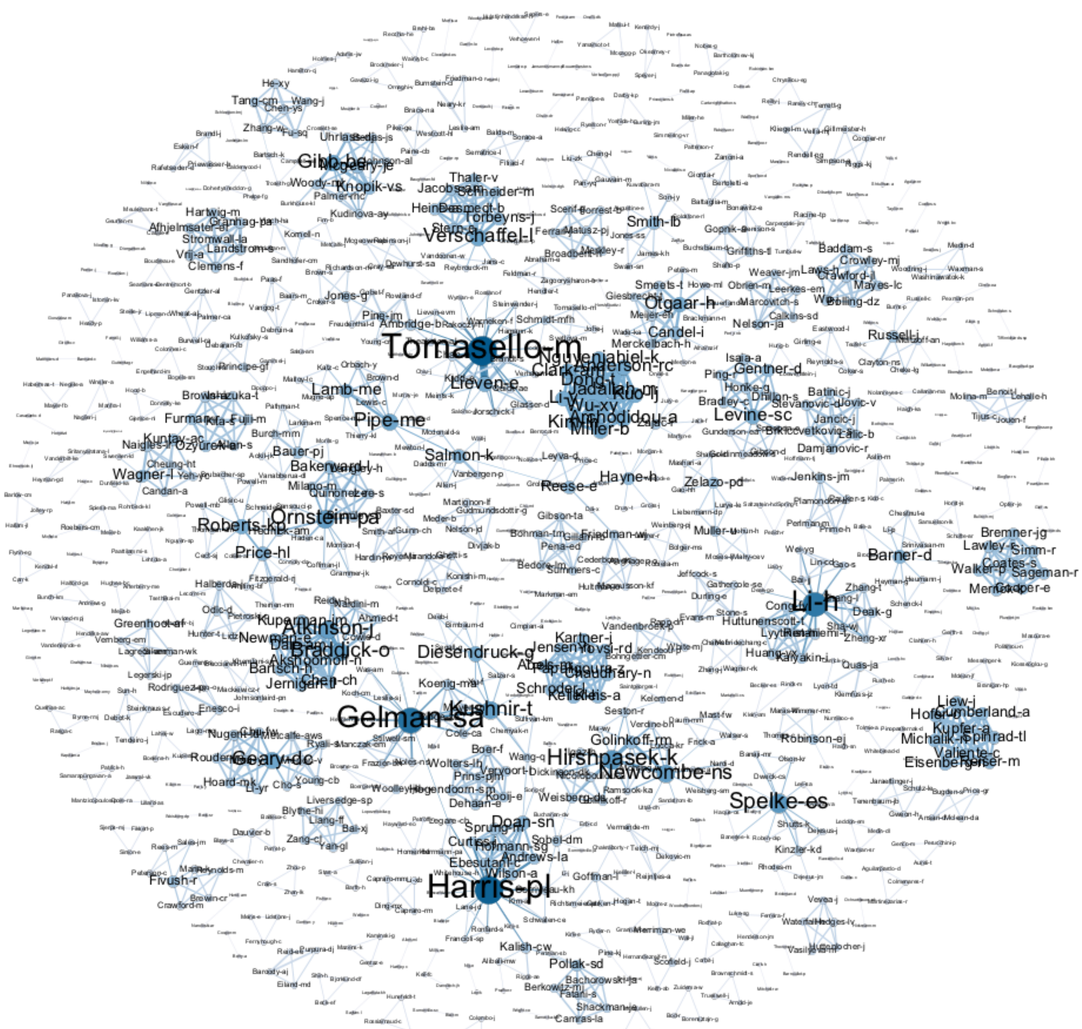
Table 5.8: Children's Coauthorship: Network Statistics

Nodes	1011
Edges	1799
Network Statistics	
Average Degree	3.628
Average Weighted Degree	3.428
Network Diameter	15
Graph Density	0.003
Modularity	0.979
Connected Components	243
Nodes	
Average Clustering Coefficient	0.912
Edges	
Average Path Length	4.612

who published in works associated with the largest connected component. For example, Tomasello-m produced thirteen publications, all in the core domain within Humanities and Psychology, but had a very specific co-authorship network that did not connect to other co-authorship clusters.

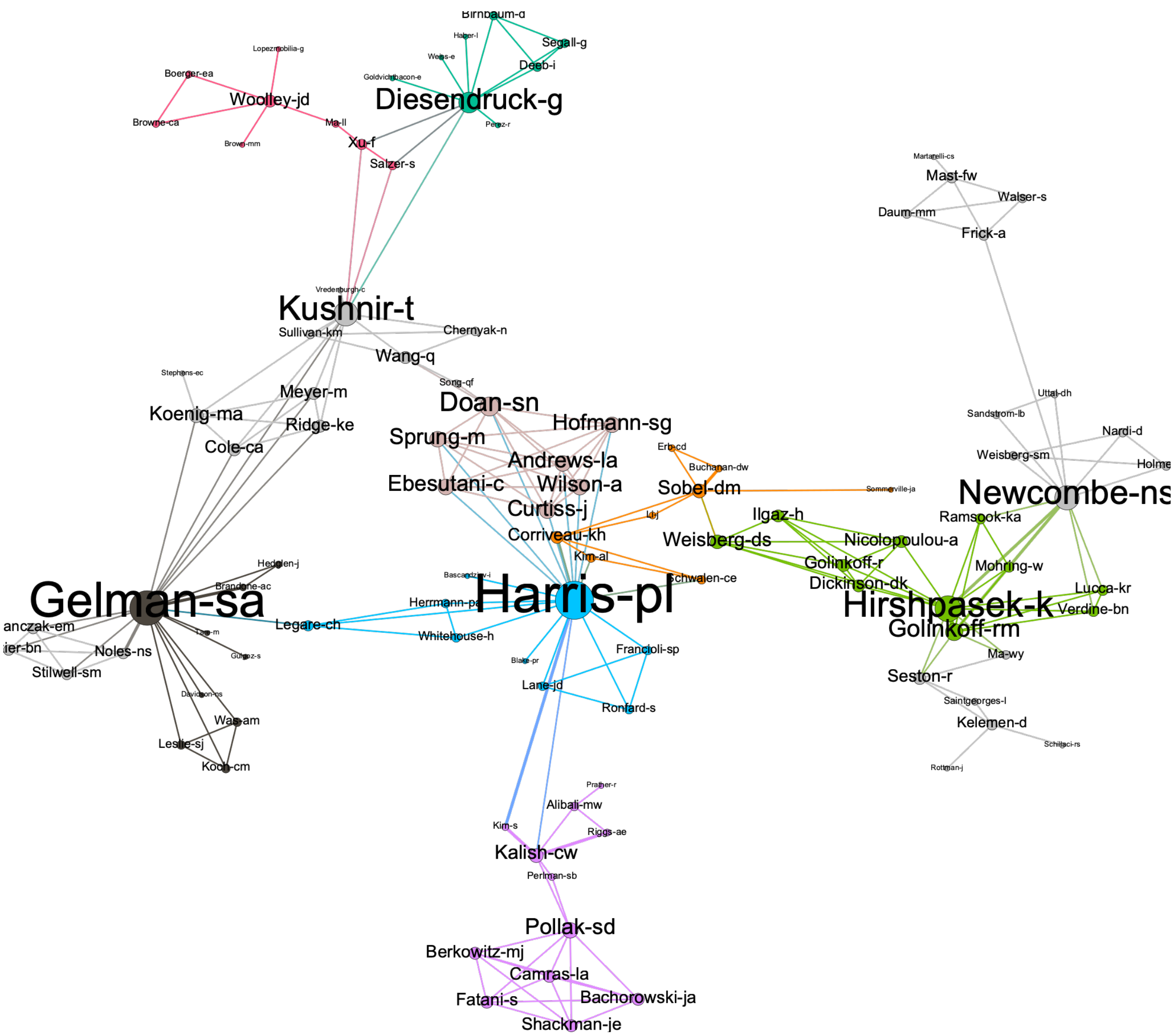
Of these, the largest connected co-authorship component consisted of 104 total authors, which can be seen in Figure 5.11. Using the default clustering algorithm in VosViewer, the 104 authors were included in a total of 14 clusters. Figure 5.11 has nodes representing authors that are colored based on cluster membership, and was visualized using the Force Atlas layout in Gephi. In this visualization, it is easy to see the authors who co-author with members in other authorship clusters, despite the algorithm defaulting to single-cluster assignment. A total of thirty-four authors, or 33.66% of authors published in both the core and extended domain. Five authors, or 4.95% of authors in the largest component, published in both the core in the discipline of Psychology and in the Extended set in Psychology and one other discipline. In clusters 2, 3, 4, 5, and 8, boundary spanning authors published in the core domain in the discipline of Psychology and in the extended domain in psychology plus one other discipline. These authors include Nicopoulou-a (Humanities), Harris-pl (Biomedical Research), Brandone-ac (Clinical Medicine), Corriveau-kh (Clinical Medicine), and Sprung-m (Clinical Medicine). Each of these authors published on the concept of "children's" in their respective secondary

Figure 5.10: Children's Co-authorship Network: All Authors



disciplines only once.

Figure 5.11: Children's Co-authorship Network: Largest Connected Component



The most central and productive authors did not necessarily engage in acts of boundary crossing. Those who did not engage in boundary spanning on the topic of children's through the act of publication include Gelman,sa (cluster 4, 25 publications), Newcombe-ns, (cluster 11, 8 publications), Diesendruck-g (cluster 7, 15 publications). Those who did included Corriveau-kh (cluster 5, 8 total publications) who published in both Psychology and Clinical Medicine; Brandone-ac (cluster 4, 4 publications), who published in both Psychology and Humanities), and Harris-pl (cluster 3, 17 publications), who published in both Clinical Medicine and Psychology. Thus, the most productive authors working on this concept did not necessarily engage in acts of boundary crossing.

5.6.3 Coauthorship: Case Study

The co-authorship network for "case study" had a network density of at 0.1, meaning the number of possible connections as opposed to the number of actual connections was very low (see Table 5.9. The average degree of connection within the co-authorship network was 2.828, and ranged from zero connections (sole authorship) to eight connections. This indicates that each node has an average of 2.828 other nodes connected to it, or in this case, other authors. The average path length, or the average number of connected nodes was 1.054, indicating smaller clusters of connections between authors publishing on this topic.

Table 5.9: Case Study: Network Statistics

Nodes	285
Edges	403
Network Statistics	
Avg. Degree	2.828
Network Diameter	2
Graph Density	0.1
Modularity	0.973
Connected Components	88
Nodes	
Avg. Clustering Coefficient	0.989
Edges	
Average Path Length	1.054

A total of 243 authors published on the concept "case study," only nine of which were connected. VosViewer detected 88 clusters, while Gephi detected 285 nodes and 403 edges. Co-authorship on this topic contained 88 weakly connected components, or clusters. Visualizing these using the OpenOrd algorithm and no filters resulted in the visualization in Figure 5.11.

Only nine authors were connected to one another through the act of coauthorship in the "case study" co-authorship analysis (Figure 5.12). These authors inhabited two clusters, shown in blue and yellow. As can be seen tin the figure, Vidal-mcr spans the boundaries of the clusters through the act of co-authorship.

No authors authoring works on "case study" engaged in acts of boundary crossing

Figure 5.12: All Case Study Co-authors

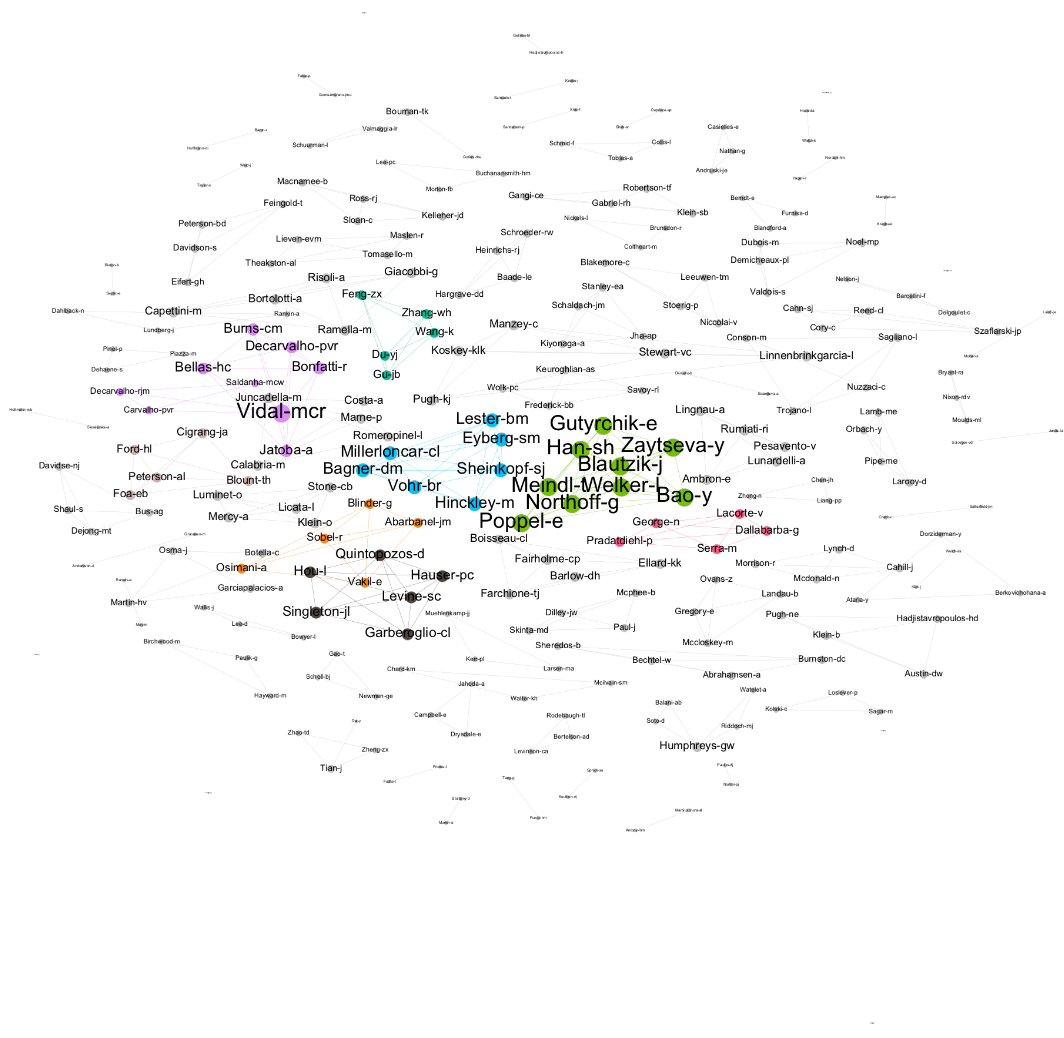
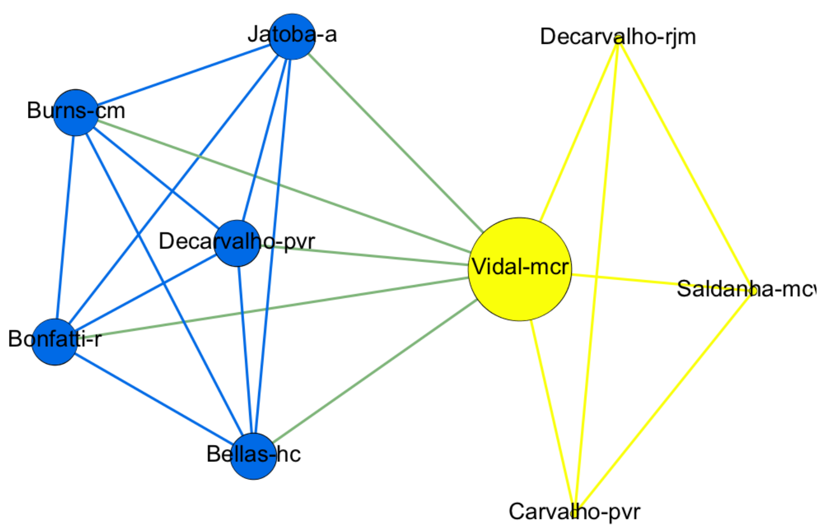


Figure 5.13: Case Study Co-authors: Largest Connected Component



through the act of publication. All nine authors who published in what VosViewer detected as the largest connected component published in the extended level of the domain in the SSH discipline of Psychology.

5.7 Conclusion

The domain of cognitive science demonstrates a growth in the diversity of concepts published on and the number of authors publishing in the domain. The seed and expand method resulted in a two-tiered analysis where the number of works published in the core continued to increase between the years 2011 and 2012, but at that point, the works cited by the core began to decrease. Based on the selection of concepts for analysis, that of the population of children's is less connected in terms of the periphera of the concepts to clusters of methodologies used to know about research problems applicable to the population. For example, "mental health service," "treatment," "implementation," and "infrastructure." Another set of particular interest pertains to gender, parenting stress, and expressive language. Authorship for the population of children's is much more connected. In part this could be due to the larger set of records related to children's (2,444) as opposed to the smaller set of records for case study (93).

Chapter 6

Discussion

6 Discussion

Analyzing the sociology of conceptual work does not guarantee that a concept-centric network in a cohesive transdisciplinary area will follow the same set of properties. It is logical to assign a potential type of classification of "subject" to the concept of "children's." It is also logical to assign "case study" to a type of study. Given that only two concepts fit the criteria for analysis of co-authorship, no solid conclusions can be drawn from network properties. However, it is possible that social networks surrounding methodology concepts have different properties than those surrounding concepts related to subjects, and thus to subjects related to specific diseases.

Had selection of concepts for analysis been limited to those that occurred with the highest frequency, not the greatest interdisciplinary reach, the resulting analysis would be quite different. Concepts with a higher frequencies of occurrence spanned different, and fewer, NSF disciplinary categories. For example, "individual differences" occurred a total of 232 times (core + extended) and was published on in every year within the core of the domain. However, this concept spans five disciplines instead of six. With no work published in the sixth discipline in the selected dataset, co-occurring concepts and authors from that discipline would not be represented. The co-occurring concepts would be more related to the five disciplines, and concepts related to "individual differences" in the sixth simply would not be present.

Both highly interdisciplinary concepts were largely contained within the NSF disci-

pline of Psychology, with children's at 91.558% and "case study" at 76.087%, but the distribution of each concept's usage in other discipline-level categories varies. 64.62% of publications in the core level were in the discipline of Psychology. This, coupled with the nature of Psychology, as a discipline involving people, makes sense that "children's" is largely discussed in Psychology journals.

A question that remains is, philosophically, what is our understanding of what makes big data "big?" A mere 4mb of natural language text is meaningless for determining the extent of the analysis unless the method (or series of methods) is taken into account in an attempt to quantify size. Complexity, or the number of steps taken in an analysis, can have any number of meanings. For example, twenty-five years of cognitive science research published between 1992 and 2016 resulted in over 1.43 million records for authors with paper data. With the progression of each year, exponential growth in the total publications (core plus extended) authors and works both year by year, citations made in each paper result in exponential growth in both the number of works published and publishing authors in the domain.

Faceted classifications, or the examination of different aspects of a subject from which to create a more holistic representation of the artifact within a system, are complex classification systems. Examples can be found throughout the 20th and 21st centuries, and include S. R. Ranganathan's Colon Classification system, the Universal Decimal Classification, and the *Getty Art and Architecture Thesaurus* (*AAT*). Getty's *AAT* is updated to include new concepts as needed. Methodology presented in this work could be used to both construct and to verify thesaural relations for interdisciplinary topics through tracing the threads of concepts and their co-occurring concepts, and verifying these against available subject classification frameworks.

Analyzing topicality changes over time has been done, but not with maintaining underlying taxonomic structure of data. In this case, the taxonomic structure was value-added NSF classification codes. Taxonomic structures in so-called "flat" file formats, such as .csv and, I would argue, most data frame structures (e.g. R's data.frame and Python's Pandas DataFrame), collapse any taxonomy into the number of levels available minus

one. For this analysis, in order to cross-reference identified boundary crossing authors, the top two levels of the taxonomy were kept. These included the Grand Discipline, and the Discipline levels. The third was too specific to glean any meaningful analysis from without analyzing a more specific topic within the oeuvre of cognitive science, such as research done on Alzheimer's Disease. Meaning, the number of categories available that produce work on the topic of Alzheimer's was too great to yield any meaningful conceptual representation of interdisciplinary concept spaces, as that presented in this work.

Large networks are inherently difficult to analyze for specific topicality. Especially in knowledge organization, work done showing the interconnectivity of authors, ideas, and places is limited. The decision to analyze the largest connected component as a starting point for examining boundary spanning, as in 5.11 was deliberate. Alternative methods of sampling were contemplated, and two were attempted. Those attempted included using VosViewer's "suggestions," and then visualizing the result in Gephi; including nodes of only a certain frequency of occurrence; and attempting to visualize from the standpoint of each of the six core disciplines in CS. The last option was only partly done, as discussing the interconnectivity of a concept from six different perspectives is its own work.

Questions concerning the extension of a domain arose during the analysis. In a seed and expand approach, the seed is traditionally the core from which the expansion flows, and encompasses the entirety of the works cited by all publications contained in the seed. The expansion, or references, included are not omitted on the basis of the time of publication. By imposing a KO-based tradition of domain analysis and relying on temporal factors, namely the year of publication, the temporal boundaries of the domain created were the result of a limit of the extension of the entirety of the sample. Thus, all works contained in the domain were published between the years of 2006 and 2016. The resulting analysis consists of two levels plus the penumbra of the meta-domain in which they reside: 1) the seed, or core of the domain, and 2) the expand, or second level of the domain.

Limiting the extension of the domain to a set of years follows a more traditional

domain-analytic paradigm, as described by Hjørland and others. Imposing this temporally constrained tradition on what is often considered a scientometric method created more manageable results for a complex multi-method analysis, as there were fewer variables of work done in the past. Each time slice analyzed for topicality for the years 2006-2016 was done on work that was published in that year, regardless of whether or not the published work was done in a "cognitive science" journal, or one that was cited by "cognitive science."

Examining patterns of authorship for boundary crossing behaviors against the most boundary spanning core concepts revealed that significantly more collaboration occurred surrounding "children's" than did surrounding "case study." This is likely due to the sense of each boundary spanning concept–children are a population, and a case study is a type of method. A population is more likely to be the focus of a researcher, and methodology follows from the nature of a research question or questions.

6.1 Assumptions

Natural language processing (NLP) tools are created operating on the assumption that a corpus contains conceptually similar enough content that it can be automatically classified as a cohesive unit. One of the shortfalls of the current NLP paradigm is that text is analyzed in a corpus; the assumptions underlying a corpus are that the documents gathered are self-similar enough in terms of topicality that features extracted from the corpus of the text will reveal an underlying theme in the "domain" of the corpus. As discussed in Chapter 1, domains can be artificially reduced to restrict the granularity of the analysis. This assumption only works with the existing natural language, and fails to accommodate known or unknown clustering concentrations of topical interests of sub-communities within larger domains as the corpus is treated as a cohesive unit.

Natural language examined in this study consists of author-created titles of publications. Titles were used for analysis because not all available articles contained abstracts, which would have resulted in an uneven analysis of topicality. In the future, this could be addressed by using a binary assignment to indicate the presence or absence of a concept

in a series of titles and abstracts, but as abstracts are more comprehensive bodies of text, the number of concepts present in a title plus an abstract would be greater than that of just an article represented by the surrogate of a title. Concepts were extracted as noun phrases using the TextBlob Python library (Loria et al. 2018). TextBlob is simple natural language processing software, and was selected in part because it allowed for the analysis of text without stemming a corpus prior to processing for noun phrases. This is a common requirement for packages such as nltk.

Work presented in this dissertation relies on the assumption that smaller communities and underlying concentrations of interest within cognitive science, clustered along the classificatory framework of the current journal-based classification paradigm, can be extracted and analyzed across boundaries present as temporal and categorical markers in the dataset. This method treats the corpus as a presumed larger domain area on the basis of natural language contained in the name of the journal title, keeping the table structures of the database intact and leveraging their value to reshape data to answer deeper questions about the social dynamics and peripheral concepts central to boundary spanning concepts. This omitted papers that may in fact be cognitive science in nature, but was a choice made to retain the classification of journals and the intentional siloing of contents, which is hidden structure of the subscriptions available of Web of Science. Another option for the selection of data would have been to consult with an expert or experts in cognitive science, and to then select journals based on their recommendations. As this was a preliminary analysis and proof-of-concept, solicited domain expertise will be incorporated in future analyses.

6.2 Web of Science Data

Data that are available through the Web of Science web interface differ from that used in this dissertation. A Web of Science subscription will return data from the web interface as a single file of up to 500 records, of which the choice of format is up to the researcher gathering data. Both .csv and the proprietary ISI format return data so that each document described is a record, and information such as authorship, ISI keywords, author

keywords, and institutional affiliations that contain multiple entries are folded into a single cell. Each individual cell containing delimiters, which indicate the further division of the data along the delimiting character (typically a "|" (pipe) or a ";" (semi-colon) are used, but tab stops are also common). Data used for this dissertation originated from a relational database. Each document described was separated with each dimension related to it into tables, using the article ID as the primary and foreign keys for each article. It must be noted that recreating this exact dataset would be extremely difficult, if not impossible, if it were to be attempted using the WoS website interface.

In order to process these data using tools readily available for general data science applications, the axes along which data were dived were the "core" and "extended," and top- and second-level NSF classifications, as well as by year. This provided a framework to subset data in R to rejoin along the article table. This data origami was necessary for analysis of the temporal nature of concepts over time, as this is not something that is typically analyzed as a bag-of-words over time.

Tools for processing Web of Science and other bibliographic database data assume that the data were gathered directly from the database. Because of this, data required reprocessing in order to accomplish certain analyses. It was necessary to restructure authorship data into a format that was usable by software. With some trial and error, different methods of restructuring data were attempted so it could be read into and processed by bibliometric software. Some software, such as VosViewer, just required the column header to match. Others, such as bibliometrix, required very specific formatting, and simply renaming the columns was not sufficient.

6.3 Moving Outside of Prepackaged Tools

Data science tools and methods allow for seemingly infinite ways to ask questions of traditionally structured data. By moving outside of pre-packaged tools for bibliographic analysis, it was possible to fracture and fold the structure of data on itself, resulting in what I will refer to as "epistemic origami." For example, a tool such as VosViewer allows for a one-step analysis of networked data. It imports the bibliometric data in a

specific file format (usually ISI or Scopus), and aligns the structures available in the file to those that are mapped in the program to each of its available analysis algorithms. This limitation restricts the types of data that can be analyzed without altering the file to fit a format the program understands. While there are very good reasons to limit the available inputs, this limits the set of questions that can be asked of data. Once everything is aligned to these structures, there is no way to further examine the data in more detail. If a researcher were to have questions other than those that could be answered by the available algorithms in the available order, the would not be able to do it using the software (without significant effort).

In constructing the methodology for this examination of a transdisciplinary area, it became apparent that most tools that currently exist for analysis of bibliometric and scientometric data are designed for much simpler analyses, and some tools did not allow for the adjustment of a visualization beyond a few options such as color, heat maps, and node views. VosViewer typically allows for one-step analysis of a single facet of structured data, and the software contains pre-packaged sampling algorithms to reduce processing time. Gephi allows for more comprehensive analysis of network data, but requires more expertise to create a visualization that conveys the connections present in the data.

6.4 Examining Concepts in Natural Language

Stemmers operate on the assumption that a set of rules can be applied to truncate words within a text to map to shorter unites that calculate to have the same base of meaning. Arguably, stemming is not the most accurate way of determining the root of meaning of a term; examples such as university/universe Manning et al. (2009) demonstrate the imperfect nature of stemming, as both terms stem to "univers," but have distinct meanings. Lemmatization is a better way of determining the best form of representation for a set of related terms, as it relies on the statistical occurrence of a given term to choose the best form. Lemmatization still has drawbacks, including that lemmas must be computed for each corpus, and requires a separate algorithm from most NLP packages. For lemmatization, the most appropriate word form is only relevant to its corresponding corpus. Of

stemming and lemmatization, stemming is the faster and more readily available method, and is included in pre-packaged bibliometric software.

Concepts occurring in the greater domain appear to represent represent anatomical structures, diseases, types of literature reviews, and the methods by which research is conducted. In the same vein, single-occurrence terms do not contain enough context to determine the topic. Despite this, single-occurrence terms are useful in determining an overall area of interest in the context of the title of the paper. As the data were regarding cognitive science, it was unsurprising that one of the top-occurring single-word terms was "brain." This in of itself is a logical result, as the brain is the center of cognition, and is thus the focus of much of cognitive science.

Adding and maintaining classificatory structures to natural language data is beneficial for many reasons. For one, the concept is given context within the realm of recorded knowledge. "Brain" can refer to a number of ideas, including the gray matter contained in the skulls of vertabrates; an individual's intellectual capacity; or to hit someone on the head with an object (paraphrased from Merriam-Webster). "Brain" in the context of Psychology takes on more meaning, as the discipline of Psychology is the study of the human mind and its functions and dysfunctions. "Human mind" takes on a more philosophical and abstract term, as opposed to the direction taken by Clinical Medicine. Thus, the imposed conceptual structure lends context to the observer for the interpretation of a concept.

Similarly, "children's" was not a complex concept. However, it met the criteria of occurring in all six core disciplines, thus making it a highly interdisicplinary concept. The surrounding periphera of concepts as present in the co-authorship network in Figure 5.8 illuminate different linkages of thought between the population of children and other phenomena and methodologies used in conjunction with the concept under examination. This suggests possibilities for combinations of analysis including sets of co-occurring concepts, such as the examination of a population (e.g. "children's"), an aspect of research (e.g. gender), and a method or methods used to learn about the linked phenomena. This tripartate examination would pinpoint more underlying epistemologies, and then using

a different methodological pipeline, where concepts in research are examined as a linked phenomenon between populations, the aspects of interest in relation to the population or primary phenomenon, and the methodology (or way of knowing–in this case, closer to Hjørland's epistemology). This would reveal how what is known is itself known.

Although not in the scope of this work, dissembling and analyzing the sociology surrounding concepts within a discipline may be a way to help validate classification systems. For example, the concept "children's" was published on in all fourteen disciplines contained in the transdisciplinary domain of cognitive science. Focusing on the social structure of a core sample of the class of things that belong to "cognitive science" revealed social clusters that aligned closely to the National Science Foundation taxonomy. A different system would, of course, yield different results. A benefit of the NSF classification is that the system is a taxonomy. By definition, a taxonomy can have one and only one space for a concept. Using a thesaurus, such as the ISO, would introduce more ambiguity, but could also clarify interdisciplinary classification problems through the use of preferred terms (PTs) and see also (SA)[1].

Examining co-authorship surrounding the concepts "children's" and "case study," and any other given concept, would result in different results depending on data sampling. Using articles on cognitive science, instead of choosing a journal-specific based approach provided these results. As citation is a social act, the transdisciplinary data likely reflects the underlying sociology surrounding authors identifying their work as belonging to cognitive science.

The choice to use a co-authorship network to visualize patterns of collaboration in a network instead of using MDS was deliberate. Co-authorship network nodes are connected by clearly visualized edges, with the underlying meaning that author A has worked with author B. MDS results visualize patterns of co-authorship as spatial phenomenon without the supporting edges indicating linkages made explicit by works co-authored by a set of authors.

[1]The author of this work acknowledges that she has no idea, at this moment, whether the ISO 9000 has been applied to journal data, and can only imagine what a massive undertaking that would be.

6.5 Concept-bearing Terms

High-frequency concept terms identified in the dataset represent anatomical structures, phenomena, diagnosable conditions and diseases, methods of study, time frames, and academic meta-descriptive terminology. The co-occurrences of concept bearing terms provide context for the use of a concept, how a concept is studied, and understanding of highly interdisciplinary terms. Children's, a concept describing either a population of study or the age range of a studied mechanism, can be broken down into research fronts and types of investigations common to research problems published in this interdisciplinary domain. This is visible in Figure 5.8 on page 93. While the categories of co-occurrences are admittedly open to interpretation, first glance shows work done on subsets of interests surrounding new research including that done on disorders in children (young person, disorder, preliminary investigation, validity), mental health services and treatment for children (mental health service, treatment, infrastructure, implementation), children's cognition, children's recall performance (working memory performance, item recall duration, insight, proactive control), and others.

Applying this same examination to case study, which is a well-known study type in which a detailed analysis is conducted for a single instance of a population or situation, either statically or in the duration of a given timeframe, the difference between a population type (children's) and a type of study (case study) emerge.

6.6 Potential Applications

Categorizing concepts into facets of research was out of the scope of this work, and will require domain expertise to meaningfully complete. However, this type of value-added analysis has enormous potential for providing more context to informetric data. Examining natural language concepts through a faceted lens would provide a clearly defined categorization of heterogenous types of concepts against to which examine mechanisms of cause and effect, types of studies commonly used for populations in a domain. This additional layer of meaning could save time by forcing the window of serendipity in information seeking for not only interdisciplinary scholars, but adding more meaning to new

concepts for students, policy makers, and the general public.

Chapter 7

Future Work

7 Future Work

Potential future directions for boundary concept analysis are many. Those discussed in this chapter include a limited set of options, and are not exclusive to the domain of cognitive science. Having examined social patterns in published literature surrounding a static window of a concept, temporal analysis of work done around a selected set of concepts key to an area of research would provide insight into the social flows surrounding different concepts. Given that "case study" is a type of method, and "children's" is roughly a population of interest, the social flows surrounding methods and populations may be drastically different. Comparing these could provide insight into what makes a topic, or an area of research, interdisciplinary. Other possible combinations of methods would have resulted in a targeted analysis of interdisciplinarity from other perspectives. Some of these include:

- starting with analyzing authorship, using an author-topic modeling approach to analyze the most interdisciplinary authors
- from analyzing highly interdisciplinary authors, examining the concepts and the conceptual content they published
- analyzing concepts in conjunction with the methodologies with which they were researched in different sub-domains of cognitive science
- examining methods and their usage within each sub-domain

The work in this dissertation was limited to titles for topical analysis. Future work addressing the uneven nature of titles in comparison of abstracts could result in the development of criterion for the inclusion of topics. Ridenour (2015) used data from the Web of Science that contained both titles and abstracts in her analysis, which resulted in the inclusion of what she termed "academic stopwords." In the development of the approach used for this dissertation, the problem of academic stopwords persisted, and was further confounded by an uneven concentration of topics when attempting to include articles in the analysis that included both titles and abstracts. An example is the inclusion of terms from an article titled "Parental Behavioral-training: An Examination of the Paradigm." High frequency term occurrences for this article are outlined in Table 7.1.

Table 7.1: Concept Concentration: Title and Abstract vs. Title

Term	Title + Abstract	Title
Paradigm	4	1
Behavioral	2	1
Conditions	2	0
Identified	2	0
Parent	2	1

Analyzing the concentration of concepts requires even representation of the extent of text, that is to say, if titles and abstracts are included in analysis for some articles, they must be included for all otherwise. Otherwise, the concentration of topics is not even. Titles have been found to contain words that represent a reasonable surrogate of the aboutness of documents O'Neill et al. (2017). As not all document records present in Web of Science contain abstracts, analyzing titles and abstracts would create an uneven picture of the overall topicality of any given domain. Taking this a step further, and focusing on topics and citations within the literature review would pull in the peripheral concepts associated with the focus of a publication. Logically, examining the methods would be the next step, and would reveal previous topics and the methods with which they were examined. This analysis of concentration would be limited to the topicality of a body of work, or corpus, only.

Analyzing journal-journal cross-disciplinary citation would aid in understanding the

intellectual structure of the lattice of journal relationships. It was intentionally omitted from this analysis, as it added unnecessary complexity. In order to accomplish this type of analysis, another method of data curation may be better suited, such using a single set of carefully selected journals for a target interdisciplinary area.

7.1 Pivot Points

From CWA, the idea of "pivot points" in terminological use examines use of official terminology as it compares to colloquial usage (Marchese and Smiraglia 2013, 255). Such pivot points may be used to provide a way to organize data from researchers to explain their conceptualization of the boundaries between information seeking, synthesizing, cross-disciplinary communication, articulating their ideas, publication, and future use of published research. Examining this idea in a single interdisciplinary area has the potential to reveal underlying assumptions and views regarding such points of discussion, and could lead to ways to classify them to aid in interdisciplinary information retrieval.

Potential pivot points include single word concept terms such as "substance," which is associated with substance abuse, substance abuse treatment, and substance abusers. The first I will discuss, substance abuse, is a narrower term of "drug abuse" in the PsycInfo Thesaurus. The term refers to the the use of drugs in both quantity and method of use that are harmful to an individual and to others. The second, "substance abuse treatment," relates to the treatment of substance abuse. The third, "substance abusers," relates to individuals who currently, or in the past, engaged in the act of substance abuse.

Tracing of concepts between areas over time was not completed in this analysis. Due to the flexible nature of the dataset, the structure required by most software to analyze this required strict data structures that were lost during the process of folding the dataset in such a way that more complex analyses were possible. As the data were already outside of most bibliographic structure, reassembling the entire dataset into a format that would allow for this type of analysis is, in of itself, its own project.

7.2 Classifying Concepts

Next steps for understanding the different conceptual facets of research present in the data could be done in a few different ways. One option would be aligning concepts to an existing thesaurus. This could also be done qualitatively, through open coding done by experts in order to create a corpus to tag the types of concepts present in noun phrases. This method would be similar to that of the Drug-Drug Interaction Corpus, which tags drugs, mechanisms, and the interactions between them. Expert input would be required to distinguish between the types of concepts, as labeling an identified conceptual phenomena a condition vs. a disease contains degrees of nuance outside of the expertise of most information scientists.

Nuances present in labeling concepts as a certain class creates implications by the inclusion of membership in that class. Creating these classes without the input of experts has the potential to create inaccurate or harmful associations for different conditions. In this case, harm is associated with the idea of othering a community, as well as the potential for psychological harm in relation to how a population is discussed.

7.3 Collocating Concepts

Next steps for understanding the periphera of related conceptual content include the use of collocations. This involves the processing of text to examine the frequencies of co-occurrences of concepts. Tracing concepts through a research problem that is the focus of a publication, the method or methods used in relation to the core of a domain, the category of each cited publication, in conjunction with a second level of cited works could be used to create a heterogeneous picture of how methods are borrowed. This particular analysis would work to perform a data-driven analysis of Bechtel's (1986) and Vinck's (2000) discussions of interdisciplinary integration. Discussing interdisciplinary integration in this way would align concepts and methods to borrowing and repurposing.

7.4 Representing the Structure of Cognitive Science

Next steps for analyzing the intellectual structure of the domain of cognitive science involve determining the best informetric method would be best adapted to categorical data. Boyack and Klavans (2010) examined thirteen different similarity-based approaches to data from the National Institute of Health. The three citation-based methods they reported on included co-citation analysis, bibliographic coupling, and direct citation. Instead of a restriction o journals and categories, articles selected were required to contain at least five MeSH terms.

To implement author co-citation analysis, expanding the temporal restrictions of the domain to a longer period of time would be appropriate. Many studies using author co-citation analysis analyze time periods greater than or equal to twenty years (Ding et al. 1999). Author co-citation analysis has long been used to illustrate the intellectual contributions of authors in an area, and is the structural basis for the visualization of the intellectual structure of different areas of research including information science (White and McCain 1998), information retrieval (Ding et al. 1999). More recently, author co-citation analysis has been applied to temporal data in knowledge management González-Valiente et al. (2019), and studies in information science comparing different ways of weighting ACA analyses (Zhao and Strotmann 2020, Bu et al. 2020) show that selecting criteria for the weighting of authorship in a work (e.g. author order, as in Bu et. Al), create different results than computing authorship equally. In this case, the selection and weighting of authors would be done on the basis of frequency of author publication, and to follow suit with the nature of boundary-crossing work, would require accommodation in the analysis for boundaries and boundary-crossing. One way this could be accomplished is by determining an author's "home" discipline, and weighting the interdisciplinarity of each author through a measure of the number of categories publication venues span.

Applications Additional applications of boundary object analysis to identify commonalities in not only different areas (e.g. policy, mental health), but different sources of data. Sources would vary on the basis of the domain selected, but include combinations of different stakeholder data. These include, but are not limited to, data-driven to empirically

investigate:

- Domains such as health-specific topics (not just oeuvres of published work), and examining the work of researchers in combination with social media data of the research population.

- Policy work examining the points of interest between produced policies and their effects on populations.

- International sets work done on important topics, such as climate change and COVID-19.

Data for many of these topics are publicly available through organizations such as the World Health Organization's Global Health Observatory (*GHO*) data (Organization 2020).

7.5 Conclusion

Analyzing an oeuvre of published work using a concept-centric approach allows for the illumination of concepts. Applying boundary object theory to a concept-centric analysis of categorized text data highlights those concepts that span one or more boundaries present as classificatory silos. This type of approach has great potential to identify niche and newly evolved areas of interdisciplinarity. It must be stressed that this work was not intended to espouse that the more interdisciplinary an object of interest is, the better the object is; instead, this work is an exploration of how ideas and those who work on these ideas intermingle in a broader information space.

References

Ahlgren, P. and B. Jarneving (2008). Bibliographic Coupling, Common Abstract Stems and Clustering: A Comparison of Two Document-Document Similarity Approaches in the Context of Science Mapping. *Scientometrics 76*, 273–290.

Ahlqvist, O. (2008, January). In search of classification that supports the dynamics of science: the FAO Land Cover Classification System and proposed modifications. *Environment and Planning B-Planning & Design 35*, 169–186. WOS:000254484500010.

Akerman, M., A. Kilpio, and T. Peltola (2010, March). Institutional change from the margins of natural resource use: The emergence of small-scale bioenergy production within industrial forestry in Finland. *Forest Policy and Economics 12*, 181–188. WOS:000275011200004.

Akkerman, S. F. and A. Bakker (2011, June). Boundary Crossing and Boundary Objects. *Review of Educational Research 81*(2), 132–169. WOS:000290399700001.

Albrechtsen, H. (1993). Subject analysis and indexing: from automated indexing to domain analysis. *Indexer 18*, 219–219.

Albrechtsen, H. and E. K. Jacob (1997). Classification systems as boundary objects in diverse information ecologies. *Advances in Classification Research Online 8*(1), 1–18.

Albrechtsen, H. and E. K. Jacob (1998). The dynamics of classification systems as boundary objects for cooperation in the electronic library. *Library Trends 47*(2), 293–312. WOS:000078568100007.

Allen, D. (2009). From boundary concept to boundary object: The practice and politics of care pathway development. *Social Science & Medicine 69*(3), 354–361.

Bateson, G. (1972). *Steps to an ecology of mind / Gregory Bateson.* New York: Ballantine Books.

Baudrillard, J. (1968). *The system of objects.* Number 3 in Radical thinkers. London: Verso.

Bechtel, W. (1986). The nature of scientific integration. In W. Bechtel (Ed.), *Integrating scientific disciplines*, pp. 3–52. Netherlands: Springer. 10.1007/978-94-010-9435-1_1.

Beghtol, C. (2003, April). Classification for Information Retrieval and Classification for Knowledge Discovery: Relationships between "Professional" and "Naive" Classifications. *Knowledge Organization 30*(2), 64–73.

Borgman, C. L. (1989). Bibliometrics and Scholarly Communication Editor's Introduction. *Communication Research 16*(5), 583–599.

Bowker, G. C. and S. L. Star (1999). *Sorting things Out: Classification and its consequences*. Cambridge, Massachusetts: MIT Press.

Boyack, K. W. and R. Klavans (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for information Science and Technology 61*(12), 2389–2404.

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

Boyack, K. W., H. Small, and R. Klavans (2013, September). Improving the accuracy of co-citation clustering using full text. *Journal of the American Society for Information Science and Technology 64*(9), 1759–1767. WOS:000322870300001.

Bradford, S. C. (1934, January). Sources of information on specific subjects. *Engineering: An Illustrated Weekly Journal 137*, 85–86.

Brewer, J. and A. Hunter (1989). *Multimethod research: A synthesis of styles.* Sage Publications, Inc.

Bu, Y., B. Wang, Z. Chinchilla-Rodríguez, C. R. Sugimoto, Y. Huang, and W.-b. Huang (2020). Considering author sequence in all-author co-citation analysis. *Information Processing & Management 57*(6), 102300.

Cabanac, G., I. Frommholz, and P. Mayr (Eds.) (2020). *Intrication between Information Retrieval and Bibliometrics: the Case of Scientific Domain Delineation.* CEUR-WS.org.

Campbell, D. T. and D. W. Fiske (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin 56*(2), 81.

Carlile, P. R. (2002, August). A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development. *Organization Science 13*(4), 442–455. INFORMS.

Carlile, P. R. (2004). Transferring, translating, and transforming: An integrative framework for managing knowledge across boundaries. *Organization science 15*(5), 555–568.

Caudill, W. and B. Roberts (1951). Pitfalls in the organization of interdisciplinary research. *Human Organization 10*(4), 12–15.

Collins, R. (1998). *The Sociology of Philosophies: A Global Theory of Intellectual Change* (First, Second Printing (2000) ed.). Cambridge, Massachusetts: The Belknap Press of Harvard University Press.

Corrales-Garay, D., M. Ortiz-de Urbina-Criado, and E.-M. Mora-Valentín (2019). Knowledge areas, themes and future research on open data: A co-word analysis. *Government Information Quarterly 36*(1), 77–87.

Courtial, J.-P. and J. Law (1989, May). A Co-Word Study of Artificial-Intelligence. *Social Studies of Science 19*(2), 301–311. WOS:A1989AB40600005.

Crane, D. (1969). Social structure in a group of scientists: A test of the "invisible college" hypothesis. *American Sociological Review 34*(4), 335–352.

Cronin, B. (1981). The need for a theory of citing. *Journal of Documentation 37*(1), 16–24.

Dahlberg, I. (2006). Knowledge organization: A new science? *Knowledge Organization 33*(1), 11–19.

Dahlberg, I. (2009). Brief Communication: Concepts and Terms–ISKO's Major Challenge. *Knowledge organization 36*, 169–177.

Darden, L. and N. Maull (1977). Interfield theories. *Philosophy of Science 44*(1), 43–64.

De Bellis, N. (2009). Maps and paradigms: Bibliographic citations at the service of the history and sociology of science. In *Bibliometrics and citation analysis*, pp. 141–179. Lanham, Md: The Scarecrow Press, Inc.

Demarest, B. and C. R. Sugimoto (2015, July). Argue, observe, assess: Measuring disciplinary identities and differences through socio-epistemic discourse. *Journal of the Association for Information Science and Technology 66*(7), 1374–1387. WOS:000355858200006.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods*. New York: McGraw Hill.

Dervin, B. (2003, October). Human studies and user studies: a call for methodological interdisciplinarity. *Information Research-an International Electronic Journal 9*(1), 166. WOS:000185916900002.

Ding, Y. (2011). Community detection: Topological vs. topical. *Journal of Informetrics 5*(4), 498–514.

Ding, Y., G. Chowdhury, and S. Foo (1999). Mapping the intellectual structure of information retrieval studies: an author co-citation analysis, 1987-1997. *Journal of information science 25*(1), 67–78.

Ding, Y., G. G. Chowdhury, and S. Foo (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information processing & management 37*(6), 817–842.

Easley, D. and J. Kleinberg (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge, United Kingdom: Cambridge University Press.

Edinburgh Language Technology Group (2015). Edinburgh Language Technology Group: Doing Stuff With Text.

Egghe, L. (1990). The duality of informetric systems with applications to the empirical laws. *Journal of information Science 16*(1), 17–27.

Egghe, L. and R. Rousseau (1990). *Introduction to informetrics: Quantitative methods in library, documentation and information science*. Elsevier Science Publishers.

Engeström, Y., R. Engeström, and M. Kärkkäinen (1995). Polycontextuality and boundary crossing in expert cognition: Learning and problem solving in complex work activities. *Learning and instruction 5*(4), 319–336.

Fang, Y. (2015). Visualizing the structure and the evolving of digital medicine: a scientometrics review. *Scientometrics 105*(1), 5–21.

Flichy, P. (2007). *Understanding technological innovation: a socio-technical approach*. Northampton, MA: Edward Elgar Publishing.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports 486*(3), 75–174.

Foster, A. and N. Ford (2003). Serendipity and information seeking: an empirical study. *Journal of Documentation 59*(3), 321–340.

Foucault, M. (1970). *The order of things : an archaeology of the human sciences*. London: Tavistock Publications.

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. Vintage.

Fujimura, J. H. (1992). Crafting Science: Standardized Packages, Boundary Objects, and Translation. In A. Pickering (Ed.), *Science as practice and culture*, pp. 168–211. Chicago: University of Chicago Press.

Fujimura, J. H. (1996). *Crafting science: A sociohistory of the quest for the genetics of cancer*. Cambridge, Massachusetts: Harvard University Press.

Galison, P. (1995). Context and constraints. In J. Z. Buchwald (Ed.), *Scientific Practice: Theories and Stories of Doing Physics*, pp. 13–41. Chicago: University of Chicago Press.

Gan, J., Q. Cai, P. Galer, D. Ma, X. Chen, J. Huang, S. Bao, and R. Luo (2019). Mapping the knowledge structure and trends of epilepsy genetics over the past decade: A co-word analysis based on medical subject headings terms. *Medicine 98*(32).

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science 178*(4060), 471–479.

Garfield, E. (1991). *Journal citation reports*. Citeseer.

Gil-Leiva, I. and A. Alonso-Arroyo (2007, June). Keywords given by authors of scientific articles in database descriptors. *Journal of the American Society for Information Science & Technology 58*(8), 1175–1187.

Glänzel, W. (2015, March). Bibliometrics-aided retrieval: where information retrieval meets scientometrics. *Scientometrics 102*(3), 2215–2222. WOS:000350337000025.

González-Valiente, C. L., M. L. Santos, and R. Arencibia-Jorge (2019). Evolution of the socio-cognitive structure of knowledge management (1986–2015): An author co-citation analysis. *Journal of Data and Information Science 4*(2), 36–55.

Greisdorf, H. (2000). Relevance: An interdisciplinary and information science perspective. *Informing Science 3*(2), 67–72.

Gumerman, G. J. and D. A. Phillips (1978). Archaeology beyond anthropology. *American Antiquity 43*(2), 184–191.

Hall, H. (2003). Borrowed theory - Applying exchange theories in information science research. *Library & Information Science Research 25*(3), 287–306. WOS:000185297700003.

Hjørland, B. (1997). *Information seeking and subject representation: an activity-theoretical approach to information science.* Number 34 in New directions in information management. Westport, Connecticut: Greenwood Press.

Hjørland, B. (2002). Domain analysis in information science: eleven approaches–traditional as well as innovative. *Journal of documentation 58*(4), 422–462.

Hjørland, B. (2003). Fundamentals of Knowledge Organization. *Knowledge Organization 30*(3), 87–111.

Hjørland, B. (2009). Concept Theory. *Journal of the American Society for Information Science and Technology 60*(8), 1519–1536.

Hjørland, B. and K. N. Pedersen (2005). A substantive theory of classification for information retrieval. *Journal of documentation 61*(5), 582–597.

Holford, W. D., M. Ebrahimi, O. Aktouf, and L. Simon (2008). Viewing Boundary "objects" as Boundary Constructions. pp. 339–339. IEEE.

Howard-Grenville, J. A. and P. R. Carlile (2006, October). The incompatibility of knowledge regimes: consequences of the material world for cross-domain work. *European Journal of Information Systems 15*(5), 473–485. WOS:000241241200008.

Hsaio, T.-M. and K.-h. Chen (2017, November). Yet another method for author co-citation analysis: A new approach based on paragraph similarity. In *80th Annual Meeting of the Association for Information Science & Technology*, Washington, DC, pp. 170–178.

Hussenot, A. and S. Missonier (2010). A deeper understanding of evolution of the role of the object in organizational process: The concept of "mediation object". *Journal of Organizational Change Management 23*(3), 269–286. WOS:000279437900005.

Husserl, E. (1950). *Cartesian Meditations: An Introduction to Phenomenology.* Boston: Kluwer Academic Publishers.

Huvila, I., T. D. Anderson, E. H. Jansen, P. McKenzie, and A. Worrall (2016). Boundary objects in information science. *Journal of the Association for Information Science and Technology 68*(8), 1807–1822.

Jacob, E. (2004). Classification and categorization: A difference that makes a difference. *Library Trends 52*(3), 515–540+668. cited By 64.

Janssens, F., J. Leta, W. Glänzel, and B. De Moor (2006, December). Towards mapping library and information science. *Special Issue on Informetrics 42*(6), 1614–1642.

Jeantet, A. and D. Vinck (1995). Mediating and commissioning objects in the sociotechnical process of product design: A conceptual approach. In D. MacLean, P. Saviotti, and D. Vinck (Eds.), *Management and new technology: Design, networks and strategies*, pp. 111–129. Bruxelles: COST/European Science Foundation.

Katz, J. S. and D. Hicks (1995). The classification of interdisciplinary journals: A new approach. In *Proceedings of the fifth international conference of the international society for scientometrics and informetrics*. Citeseer.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation 14*(1), 10–25.

Klein, J. T. (1990). *Interdisciplinarity: History, theory, and practice.* Detroit, MI: Wayne State University Press.

Klein, J. T. (2000). A conceptual vocabulary of interdisciplinary science. In P. Weingart and N. Stehr (Eds.), *Practising interdisciplinarity*, pp. 3–24. Buffalo: University of Toronto Press.

Klein, J. T. (2010). *The Oxford Handbook of Interdisciplinarity*, Chapter A taxonomy of interdisciplinarity, pp. 15–30. Oxford University Press.

Kleinberg, J. (2003). Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery 7*(4), 373–397.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lamont, M. (1992). *Money, morals, and manners: The culture of the French and the American upper-middle class*. University of Chicago Press.

Lamont, M. and V. Molnár (2002). The study of boundaries in the social sciences. *Annual review of sociology 281*(1), 167–195.

Larivière, V., C. R. Sugimoto, and B. Cronin (2012, May). A bibliometric chronicling of library and information science's first hundred years. *Journal of the American Society for Information Science and Technology 63*(5), 997–1016. WOS:000303500300010.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, Massachusetts: Harvard university press.

Law, J., S. Bauin, J.-P. Courtial, and J. Whittaker (1988). Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. *Scientometrics 14*(3-4), 251–264.

Lawrence, P. R. and J. W. Lorsch (1967). Differentiation and integration in complex organizations. *Administrative science quarterly 12*(1), 1–47.

Leydesdorff, L. (2006). Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? *Journal of the American Society for Information Science and Technology 57*(5), 601–613.

Leydesdorff, L. and I. Rafols (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology 60*(2), 348–362.

Loria, S., P. Keen, R. Yankovksky, D. Karesh, E. Dempsey, W. Childs, J. Schnurr, A. Qalieh, L. Ragarsson, J. Coe, A. L. Calvo, N. Kulshretha, J. Eslava, T. J. Harden, J. Kolb, D. Ong, and J. Moschella (2018). Textblob: Simplified text processing.

Madsen, R. E., D. Kauchak, and C. Elkan (2005). Modeling word burstiness using the Dirichlet distribution. In *Proceedings of the 22nd international conference on machine learning*, pp. 545–552. Association for Computing Machinery.

Maghrebi, M., A. Abbasi, S. Amiri, R. Monsefi, and A. Harati (2011, January). A collective and abridged lexical query for delineation of nanotechnology publications. *Scientometrics 86*(1), 15–25. WOS:000284888600002.

Manning, C. D., P. Raghavan, and H. Schütze (2009). Introduction to Information Retrieval.

Marchese, C. and R. P. Smiraglia (2013). Boundary Objects: CWA, an HR Firm, and Emergent Vocabulary. *Knowledge Organization 40*(4), 254–259. WOS:000323397700004.

Maurer, M. B. and S. Shakeri (2016, June). Disciplinary Differences: LCSH and Keyword Assignment for ETDs from Different Disciplines. *Cataloging & Classification Quarterly 54*(4), 213–243.

Mayr, P. and A. Scharnhorst (2015, March). Scientometrics and information retrieval: weak-links revitalized. *Scientometrics 102*(3), 2193–2199. WOS:000350337000023.

Milojević, S. (2009). *Big science, nano science?: Mapping the evolution and socio-cognitive structure of nanoscience/nanotechnology using mixed methods.* Ph.D., University of California, Los Angeles, United States – California.

Moed, H. F. (2015). The Future of Research Evaluation Rests With an Intelligent Combination of Advanced Metrics and Transparent Peer Review. In B. Cronin and C. R. Sugimoto (Eds.), *Scholarly Metrics Under the Microscope: From Citation Analysis to Academic Auditing*, ASIS&T monograph series, pp. 693–712. Medford, NJ: Information Today, Inc.

Mohammadi, E. and M. Thelwall (2014). Mendeley readership altmetrics for the social sciences and humanities: Research evaluation and knowledge flows. *Journal of the Association for Information Science and Technology 65*(8), 1627–1638.

Narin, F., M. Carpenter, and N. C. Berlt (1972). Interrelationships of scientific journals. *Journal of the American Society for Information Science 23*(5), 323–331.

O'Neill, E. T., K. A. Kammerer, and R. Bennett (2017, October). The aboutness of words. *Journal of the Association for Information Science & Technology 68*(10), 2471–2483.

Organization, W. H. (2020). Global health observatory: The data repository.

Otlet, P. (1990). *International Organisation and Dissemination of Knowledge: Selected Essays of Paul Otlet*. FID publication.

Palmer, C. L. (1996). Information work at the boundaries of science: Linking library services to research practices. *Library Trends 45*(2), 165–191. WOS:A1996VV90300004.

Palmer, C. L. and L. J. Neumann (2002, January). The information work of interdisciplinary humanities scholars: Exploration and translation. *Library Quarterly 72*(1), 85–117. WOS:000173831500004.

Papadimitriou, K. and C. Pellegrin (2007). Dynamics of a project through Intermediary Objects of Design (IODs): A sensemaking perspective. *International Journal of Project Management 25*(5), 437–445.

Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park: SAGE Publications, inc.

Paul, M. J. and R. Girju (2009). Topic Modeling of Research Fields: An Interdisciplinary Perspective. In *The International Conference on Recent Advances in Natural Language Processing (RANLP)*, pp. 337–342.

Peirce, C. S. (1992). The essential Peirce: Selected philosophical writings. Volume 2 (1893-1913). Bloomington: Indiana University Press.

Pierce, S. J. (1999). Boundary crossing in research literatures as a measure of interdisciplinary information transfer. *Journal of the American Society for Information Science 50*, 271–279.

Pieters, R. and H. Baumgartner (2002). Who talks to whom? Intra-and interdisciplinary communication of economics journals. *Journal of Economic Literature XL*, 483–509.

Popper, K. (1972). *Objective knowledge; an evolutionary approach.* Oxford: Clarendon Press.

Porter, A. and I. Rafols (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics 81*(3), 719–745.

Postholm, M. B. (2015, January). Methodologies in Cultural-Historical Activity Theory: The example of school-based development. *Educational Research 57*(1), 43–58. WOS:000346849800003.

Price, D. J. d. S. (1963). *Little science, big science... and beyond.* Columbia University Press New York.

Price, D. J. d. S. (1965a). Networks of scientific papers. *Science 149*(3683), 510. cited By 919.

Price, D. J. d. S. (1965b). Regular Patterns in the Organization of Science. *Organon 2*, 243–248.

Reagans, R. and B. McEvily (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative science quarterly 48*(2), 240–267.

Rhoades, S. A. (1993). The Herfindahl-Hirschman Index. *Fed. Res. Bull. 79*, 188.

Ridenour, L. (2015). Concept term repurposing: Framing shifts in domains and terminology. In P. Kielty (Ed.), *2015 Noth American Symposium on Knowledge Organization*, Los Angeles, CA.

Ridenour, L. (2016). Boundary Objects: Measuring Gaps and Overlap Between Research Areas. *Knowledge Organization 43*(1), 44–55.

Ridenour, L. (2020). csv2bibtex script. Github.

Sainte-Marie, M., L. Ridenour, and V. Larivière (2018, November). Examining Communities in the Transdisciplinary Area of Cognitive Science: Automatic Classification for Examining Communities in the Web of Science Using Unsupervised Clustering Methods. In *Culture, Community, and Voice in Knowledge Organization Systems. Workshop.*, Vancouver, BC.

Salton, G. (1975). *A Theory of Indexing*. Philadelphia: Society for Industrial and Applied Mathematics.

Salton, G. (1989). *Automatic Indexing*. Addison-Wesley Series in Computer Science. New York: Addison-Wesley Publishing Company.

Sampalli, T., M. Shepard, J. Duffy, and R. Fox (2010). An evaluation of SNOMED CT® in the domain of complex chronic conditions. *International journal of integrated care 10*, 1–10.

Saussure, F. d. (1959). *Course in General Linguistics*. New York: McGraw Hill.

Shannon, C. E. (1948). A Mathematical Theory of Communication, Part I. *Bell Systems Technical Journal 27*(3), 379–423.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for information Science 24*(4), 265–269.

Smiraglia, R. P. (2008). Knowledge organization in search of a concept.

Smiraglia, R. P. (2014). *The elements of knowledge organization*. New York: Springer International Publishing.

Smiraglia, R. P. (2015, February). Keywords Redux–An Editorial. *Knowledge Organization 42*(1), 3–7.

Smiraglia, R. P. (2017). Disciplinary, Asynthetic, Domain-Dependent: NARCIS a National Research Classification in Isolation. In *Proceedings of the SIG/CR Workshop:*

Building a Research Agenda for Multi-Perspective Knowledge Representation, Washington, DC.

Smiraglia, R. P. and H.-l. Lee (2012). *Cultural Frames of Knowledge*. Ergon.

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation 28*(1), 11–21.

Star, S. L. (1988). Introduction: The Sociology of Science and Technology. *Social Problems 35*(3), 197–205.

Star, S. L. (2010). This is not a boundary object: Reflections on the origin of a concept. *Science, Technology & Human Values 35*(5), 601–617.

Star, S. L. and J. R. Griesemer (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science 19*(3), 387–420.

Strehlow, R. (1993). Terminology Standardization in the Physical Sciences. In H. B. Sonneveld and K. L. Loening (Eds.), *Terminology: applications in interdisciplinary communication*, pp. 127–140. John Benjamins Publishing.

Sugimoto, C. R., D. Li, T. G. Russell, S. C. Finlay, and Y. Ding (2011, January). The Shifting Sands of Disciplinary Development: Analyzing North American Library and Information Science Dissertations Using Latent Dirichlet Allocation. *Journal of the American Society for Information Science and Technology 62*(1), 185–204. WOS:000285428000015.

Sugimoto, C. R. and S. Weingart (2015). The kaleidoscope of disciplinarity. *Journal of Documentation 71*(4), 775–794. WOS:000359056200009.

Suppe, F. (1989). Classification. In E. Barnouw (Ed.), *International Encyclopedia of Communications*, Volume 1, pp. 292–96. Oxford, UK: Oxford University Press.

Szostak, R. (2004). *Classifying science: Phenomena, data, theory, method, practice.* Information Science and Knowledge Management. Norwell, MA: Springer.

Szostak, R. b. (2008). Classification, interdisciplinarity, and the study of science. *Journal of Documentation 64*(3), 319–332. cited By 33.

Taylor, A. G. and D. N. Joudrey (2009). *The Organization of Information* (Third ed.). Library and Information Science Text Series. Westport, Connecticut: Libraries Unlimited.

Tennis, J. T. (2002). Subject ontogeny: subject access through time and the dimensionality of classification. *Advances in Knowledge Organization 8*, 54–59.

Tennis, J. T. (2003). Two axes of domains for domain analysis. *Knowledge Organization 30*(3-4), 191–195. cited By 27.

Thomson Reuters (2013). 2013 Journal Citation Reports® Social Sciences Edition.

Tijssen, R. J. and A. F. Van Raan (1994). Mapping Changes in Science and Technology Bibliometric Co-Occurrence Analysis of the R&D Literature. *Evaluation Review 18*(1), 98–115.

Timmermans, S. (2015). Introduction: Working with Leigh Star. In G. C. Bowker, S. Timmermans, A. E. Clarke, and E. Balka (Eds.), *Boundary Objects and Beyond: Working with Leigh Star*. Cambridge, MA: MIT Press.

Trompette, P. and D. Vinck (2009). Revisiting the notion of boundary object. In *Revue d'anthropologie des connaissances*. SAC.

Tsui, A. B. M. and D. Y. K. Law (2007). Learning as boundary-crossing in school–university partnership. *Teaching and teacher education 23*(8), 1289–1301.

Uyar, A., M. Kılıç, and M. A. Koseoglu (2020). Exploring the conceptual structure of the auditing discipline through co-word analysis: An international perspective. *International Journal of Auditing 24*(1), 53–72.

van den Huevel, C. and R. P. Smiraglia (2013). Visualizing knowledge interaction in the multiverse of knowledge. In A. Slavic, A. Akdag Salah, and S. Davies (Eds.), *Proceedings of the International UDC Seminar 2013*, Wurzburg, pp. 59–72. Ergon Verlag.

Vinck, D. (2000). Pratiques de l'Interdisciplinarite: Mutations des Sciences, de l'Industrie et l'Enseignement. In *Pratiques de l'interdisciplinarité: Mutations des sciences, de l'industrie et de l'enseignement*, Genie Industriel, pp. 99–112. Presses Universitaires Grenoble.

Wagner, C. S., J. D. Roessner, K. Bobb, J. T. Klein, K. W. Boyack, J. Keyton, I. Rafols, and K. Börner (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics 5*(1), 14–26.

Weick, K. E. (1988). Enacted sensemaking in crisis situations. *Journal of Management Studies 25*(4), 305–317.

Wertsch, J. V. (1993). *Voices of the mind: A sociocultural approach to mediated action.* Cambridge, Massachusetts: Harvard University Press.

White, H. D. and K. W. McCain (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American society for information science 49*(4), 327–355.

Williams, T. (2005). Assessing and moving on from the dominant project management discourse in the light of project overruns. *IEEE Transactions on Engineering Management 52*(4), 497–508.

Wilson, P. (1993). Communication efficiency in research and development. *Journal of the American Society for Information Science 44*, 376–382.

Wolfram, D. (1992). Applying informetric characteristics of databases to IR system file design, Part II: simulation comparisons. *Information processing & management 28*(1), 135–151.

Wolfram, D. (2000). Applications of informetrics to information retrieval research. *Informing Science 3*(2), 77–82.

Wolfram, D. (2003). *Applied Informetrics for Information Retrieval Research*. Number 36 in New Directions in Information Management. Westport, Connecticut: Greenwood Publishing Group.

Wolfram, D. (2015, March). The symbiotic relationship between information retrieval and informetrics. *Scientometrics 102*(3), 2201–2214.

Yan, E. and Y. Ding (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other. *Journal of the American Society for Information Science and Technology 63*(7), 1313–1326.

Yeo, J. and S. C. Tan (2014, September). Redesigning problem-based learning in the knowledge creation paradigm for school science learning. *Instructional Science 42*(5), 747–775. WOS:000340093400004.

Zhang, G. and E. K. Jacob (2013). Understanding boundaries: physical, epistemological and virtual dimensions. *Information Research-an International Electronic Journal 18*(3), C21. WOS:000329310400030.

Zhang, J., Q. Yu, F. Zheng, C. Long, Z. Lu, and Z. Duan (2016). Comparing keywords plus of wos and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science and Technology 67*(4), 967–972.

Zhao, D. and A. Strotmann (2020). Telescopic and panoramic views of library and information science research 2011–2018: a comparison of four weighting schemes for author co-citation analysis. *Scientometrics 124*, 255–270.

Zitt, M. (2015, March). Meso-level retrieval: IR-bibliometrics interplay and hybrid citation-words methods in scientific fields delineation. *Scientometrics 102*(3), 2223–2245. WOS:000350337000026.

Zitt, M. and E. Bassecoulard (2006, December). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information processing & management 42*(6), 1513–1531.

7.6 Appendix A: Code

Script to process natural language from titles and identify and align boundary concepts across categories:

```
import textblob
from textblob import TextBlob
from textblob import Word

import sys
from importlib import reload
reload(sys)
sys.getdefaultencoding()

grdiscnse ='~/data/titles/nsecoretitles2016.txt'
file=open(grdiscnse)
t=file.read()
print(type(t))
print(type(grdiscnse))
grdiscnse = TextBlob(t)

grdiscssh ='~/data/titles/sshcoretitles2016.txt'
file=open(grdiscssh)
t2=file.read()
print(type(t2))
grdiscssh = TextBlob(t2)
```

```
titlessh = grdiscssh.noun_phrases #saves noun phrases to a
    variable
titlense = grdiscnse.noun_phrases

from collections import Counter
#make the dictionaries using Counter
dictionaryssh = Counter(titlessh)
dictionarynse = Counter(titlense)

len(dictionaryssh) #verify the length of nse
len(dictionarynse) #verify the length of ssh

for value, count in dictionarynse.most_common():
    print(value,'',",count) #prints sorted list of most common
        noun phrases for nse

for value, count in dictionaryssh.most_common():
    print(value,'',", count) #prints sorted list of most common
    noun phrases for ssh

with open("~/data/dataprocessingfiles/nse2016.txt", "w") as f:
    #for t in testnp:
        # f.write(str(t) + "\n") #this prints out the noun
            phrases as detected by textblob.
    for value, count in dictionarynse.most_common():
        f.write(str(value) + ";" + str(count) + "\n") #loops
            over and saves the nse frequencies

with open("~core/dataprocessingfiles/ssh2016.txt", "w") as f:
```

```
    #for t in testnp:
        # f.write(str(t) + ''\n") #this prints out the noun
            phrases as detected by textblob.
    for value, count in dictionaryssh.most_common():
        f.write(str(value) + '';"␣+␣str(count)␣+␣''\n") #loops
            over and saves the ssh counts

###ALIGN BOUNDARY CONCEPTS USING A PANDAS DATAFRAME###
import pandas
import numpy
from pandas import DataFrame
import os
import csv

nsecount = pandas.read_table(''~/data/dataprocessingfiles/
    nse2016.txt",␣sep=r"\;",␣engine='python',␣quoting=csv.
    QUOTE_NONE,␣names␣=␣(''NP",''count"))
#read␣in␣output␣file␣as␣a␣table␣from␣the␣previous␣output␣for␣nse

sshcount␣=␣pandas.read_table(''~/data/dataprocessingfiles/
    ssh2016.txt", sep=r"\;", engine='python',␣quoting=csv.
    QUOTE_NONE,␣names␣=␣(''NP",''count"))
#read␣in␣output␣file␣as␣a␣table␣from␣the␣previous␣output␣for␣ssh

result␣=␣pandas.merge(nsecount,
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣sshcount[['NP', 'count']],
␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣␣on␣=␣'NP')
#merges the two DataFrames, functioning as a database-style
    inner join that keeps only the nps occurring in both
```

```
    DataFrames at least once

resultpyear = pandas.DataFrame(result)
print(result)
#sets a new DataFrame equal to the result of the join

resultpyear['year'] = 2016
#adds the relevant year as a new column to the DataFrame

resultpyear.to_csv(''~/data/comparison/2016comparison.csv")
#saves the output to a .csv
```

7.7 Appendix B: Tables

Table 7.2: Level, High-level Discipline, Discipline, and Article Counts

Level	GrDisc	Disc	Num. Records
core	natural sciences and engineering	biology	1211
core	natural sciences and engineering	clinical medicine	13769
core	natural sciences and engineering	engineering and technology	556
core	social sciences and humanities	humanities	1017
core	social sciences and humanities	professional fields	359
core	social sciences and humanities	psychology	30885
extended	natural sciences and engineering	biology	4247
extended	natural sciences and engineering	biomedical research	21243
extended	natural sciences and engineering	chemistry	179
extended	natural sciences and engineering	clinical medicine	122191
extended	natural sciences and engineering	earth and space	317
extended	natural sciences and engineering	engineering and technology	7305
extended	natural sciences and engineering	mathematics	1512
extended	natural sciences and engineering	physics	2855
extended	social sciences and humanities	arts	224
extended	social sciences and humanities	health	9931
extended	social sciences and humanities	humanities	4716
extended	social sciences and humanities	professional fields	10758
extended	social sciences and humanities	psychology	97234
extended	social sciences and humanities	social sciences	6526

Table 7.3: Full Category Names, Recoded Data, Num. Records, and Num. Abstracts

Articles	Recoding	Category Name	Total Records	Num. Abstracts	% with Abstracts
(All)	N/A	(All)	337076	325108	96.45%
Level	1	Core	47797	46109	96.47%
	2	Extended	289238	279000	96.46%
GrDisc	1	Natural Sciences and Engineering	175385	169771	96.80%
	2	Social Sciences and Humanities	161650	155338	96.10%
Discipline	1	Arts	224	133	59.38%
	2	Biology	5458	5250	96.19%
	3	Biomedical Research	21243	20270	95.42%
	4	Chemistry	179	152	84.92%
	5	Clinical Medicine	135960	132111	97.17%
	6	Earth and Space	317	300	94.64%
	7	Engineering and Technology	7861	7489	95.27%
	8	Health	9931	9654	97.21%
	9	Humanities	5733	4577	79.84%
	10	Mathematics	1512	1463	96.76%
	11	Physics	2855	2743	96.08%
	12	Professional Fields	11117	10582	95.19%
	13	Psychology	128119	124442	97.13%
	14	Social Sciences	6526	5955	91.25%

Table 7.4: Children's Coauthorship by Cluster Aligned to NSF Taxonomy

		Core	Extended			
		SSH	NSE		SSH	
Cluster	Label	Psychology	Bio. Res.	Clin. Med.	Humanities	Psychology
1	Alibali-mw	2				1
	Bachorowski-ja	1				
	Berkowitz-mj	1				
	Camras-la	1				
	Fatani-s	1				
	Kalish-cw	5				2
	Kim-s	3				
	Perlman-sb	1				
	Pollak-sd	2				1
	Prather-r	1				
	Riggs-ae	2				2
	Shackman-je	1				
2	Dickinson-dk	1				
	Golinkoff-r	1				
	Golinkoff-rm	3				2
	Hirshpasek-k	4				2
	Ilgaz-h	1				
	Lucca-kr	1				
	Mohring-w	1				
	Nicolopoulou-a	1			1	
	Ramsook-ka	1				
	Verdine-bn	1				
	Weisberg-ds	2				
3	Bascandziev-i	1				
	Blake-pr	1				
	Harris-pl	9	1			7
	Herrmann-pa	1				
	Lane-jd	1				6
	Legare-ch	2				2
	Ronfard-s	1				
	Whitehouse-h	1				
4	Brandone-ac	2		1		1
	Davidson-ns	1				
	Gelman-sa	10				15
	Gulgoz-s	1				
	Hedglen-j	1				
	Koch-cm	1				
	Leslie-sj	1				
	Tare-m	1				
	Was-am	1				
5	Buchanan-dw	2				
	Corriveau-kh	3	1			4
	Erb-cd	1				
	Kim-al	1				
	Li-j	1				
	Schwalen-ce	1				
	Sobel-dm	7				6
	Sommerville-ja	1				2
6	Boerger-ea	1				1
	Brown-mm	1				
	Browne-ca	1				
	Lopezmobilia-g	1				
	Ma-ll	2				3
	Salzer-s	1				
	Woolley-jd	4				2
	Xu-f	2				3
7	Birnbaum-d	1				2
	Deeb-i	1				2
	Diesendruck-g	6				9
	Goldvichtbacon-e	1				
	Haber-l	1				
	Perez-r	1				
	Segall-g	1				2
	Weiss-e	1				
8	Andrews-la	1				
	Curtiss-j	1				
	Doan-sn	2				
	Ebesutani-c	1				
	Hofmann-sg	1				
	Sprung-m	1		1		1
	Wilson-a	1				
9	Chernyak-n	1				
	Kushnir-t	4				
	Song-qf	1				
	Sullivan-km	1				
	Vredenburgh-c	1				
	Wang-q	2				
10	Kelemen-d	3				2
	Ma-wy	1				
	Rottman-j	1				
	Saintgeorges-l	1				
	Schillaci-rs	1				
	Seston-r	2				
11	Holmes-ca	1				
	Nardi-d	1				
	Newcombe-ns	5				3
	Sandstrom-lb	1				
	Uttal-dh	1				1
	Weisberg-sm	1				
12	Daum-mm	1				2
	Frick-a	2				1
	Martarelli-cs	1				1
	Mast-fw	2				1
	Walser-s	1				

7.4 (continued)

		Core	Extended			
		SSH	NSE		SSH	
Cluster	Label	Psychology	Bio. Res.	Clin. Med.	Humanities	Psychology
13	Cole-ca	1				
	Koenig-ma	2				2
	Meyer-m	1				
	Ridge-ke	1				
14	Frazier-bn	1				1
	Manczak-em	1				1
	Noles-ns	2				
	Stilwell-sm	1				

CURRICULUM VITAE

Laura E. Ridenour

Doctoral Candidate

University of Wisconsin-Milwaukee

EDUCATION

University of Wisconsin-Milwaukee, Milwaukee, WI 2020

Ph.D, Candidate, Information Studies Major: Information Organization

Dissertation Title: "Examining the Notion of the Boundary Object in Information Systems: The Transdisciplinary Oeuvre of Cognitive Science"

Indiana University Bloomington, Bloomington, IN 2013

Master of Information Science

Portland State University, Portland, OR 2009

Bachelor of Music, Music Performance; Minor: Business Administration

PUBLICATIONS

PEER-REVIEWED JOURNAL ARTICLES

Martínez Ávila, D., Smiraglia, R., Szostak, R., Scharnhorst, A., Beek, W., Siebes, R., Ridenour, L., & Schlais, V. (2018). CLASSIFYING THE LOD CLOUD. Brazilian Journal of Information Science: Research Trends, 12(4), 06-10. https://doi.org/10.36311/1981-1640.2018.v12n4.02.p6

Ridenour, L. & Jeong, W. (2016). Are We There Yet? Calculating Wait Time For Popular Digital Titles: A Case Study. Public Library Quarterly 35, 131-156. https://doi.org/10.1080/01616846.2016.1163967

Ridenour, L. (2016). Boundary Objects: Measuring Gaps And Overlap Between Research

Areas. Knowledge Organization. 43, 44-55. https://www.isko.org/ko431toc.pdf

PEER-REVIEWED CONFERENCE PAPERS

Sainte-Marie, M., Ridenour, L., & Larivière, V. (2018). Examining Communities in the Transdisciplinary Area of Cognitive Science: Automatic Classification for Examining Communities in the Web of Science Using Unsupervised Clustering Methods. *ASIS&T Special Interest Group/Classification Research (SIG/CR) Workshop on Culture, Community, and Voice in Knowledge Organization Systems*, Vancouver, B.C. http://dx.doi.org/10.7152/acro.v29i1.15463

Martínez-Ávila, D., Smiraglia, R. P., Szostak, R., Scharnhorst, A., Beek, W., Siebes, R., Ridenour, L., & V. Schlais. (2018). Classifying the LOD Cloud: Digging Into the Knowledge Graph. X Encontro Internacional De Informação, Conhecimento E Ação (X EIICA). Marília, Brazil.

Ridenour, L. (2017). Visualizing Data, Information, and Knowledge: A Discussion. North American Symposium on Knowledge Organization, Champaign, IL.

Ridenour, L. (2016). Practical Applications of Citation Analysis to Examine Interdisciplinary Knowledge. ISKO. Rio de Janiero, Brazil.

Ridenour, L. & Smiraglia, R. P. (2016). How Interdisciplinary is Knowledge Organization? An Epistemological View of Knowledge Organization as a Domain. ISKO. Rio de Janiero, Brazil.

Ridenour, L. & Jeong, W. (2016). Leveraging The Power Of Social Reading And Big Data: An Analysis Of Co-Read Clusters Of Books On Goodreads. iConference, Philadelphia, PA. http://hdl.handle.net/2142/89314

Jeong, W. & Ridenour, L. (2016). Visualization Of Co-Read Book Data At Goodreads.Com: A Potential Readers' Advisory Tool. ALISE 2016 Annual Conference, Boston, MA.

Jeong W. & Ridenour, L. (2016). Fostering Diversity In Library And Information Science

Education: The FEAL Grant. ALISE 2016 Annual Conference, Boston, MA.

Ridenour, L. (2015). Concept Term Repurposing: Framing Shifts In Domains And Terminology. North American Symposium on Knowledge Organization, Los Angeles, CA.

PEER-REVIEWED POSTERS

Ridenour, L. (2019). Examining the Notion of the Boundary Object in Information Systems: The Transdisciplinary Oeuvre of Cognitive Science. ALISE 2019 Annual Conference, Knoxville, TN. http://hdl.handle.net/2142/105284

Ridenour, L. (2019). Examining Concepts Across Boundaries: A Pre-Processing Framework. Proceedings of the 2019 North American Symposium on Knowledge Organization. Philadelphia, PA. June 13, 2019.

Ridenour, L., Porter, D., & Emery, D. (2019). Updating Medieval Manuscript Data Using Supervised Classification: A Preliminary Process. Proceedings of the 2019 iConference, Washington, D.C. March 15, 2019. http://hdl.handle.net/2142/103349

Choi, I. & Ridenour, L. (2016). Cross-Lingual Tag Analysis of English-Tagged Literature Written In English And Korean. WAAL, Green Lake, WI. April 21, 2016.

Jeong, W., Han, H., and Ridenour L. (2015). Case Study of Waiting List on WPLC Digital Library. JCDL, Knoxville, TN. June 21-25, 2015.

Ridenour, L. & Park, H. (2014). Visual Analysis of Historical Patterns: ISKO Proceedings Titles from 1992 to 2012. ISKO. Kraków, Poland. May 19-22, 2014.

EDITORIALS

Alhumaidan, Y., Dowell, M., Rene, J., Leverett, L., Ridenour, L., Schlais, V., & Smiraglia, R. P. (2018). Knowledge Organization and the 2017 UDC Seminar: An Editorial. Knowledge Organization, 45, 273-280. doi:10.5771/0943-7444-2018-4-273.

Beak, J., Choi, I., Lee, S., Park, H., Ridenour, L., & Smiraglia, R. P. (2014). Knowledge

Organization and the 2013 UDC Seminar: An Editorial. Knowledge Organization 41, 191-194.

CONFERENCE AND WORKSHOP PROCEEDINGS

Dobreski, B. & Ridenour, L. (eds.). (2019). Proceedings from the North American Symposium on Knowledge Organization. Vol. 7. Philadelphia, PA.

Montoya, R. D., Hajibayova, L., Matsuda, S., Ridenour, L., & Zhitomirsky-Geffet, M. (eds.). (2018). Culture, Community, and Voice in Knowledge Organization Systems. Advances in Classification Research Online. Vancouver, B.C. November 10, 2018.

Hajibayova, L., Kwaśnik, B. H., Montoya, R. D., Ridenour, L., & Zhitomirsky-Geffet, M. (eds.). (2017). Building a Research Agenda for Multi-Perspective Knowledge Representation. Advances in Classification Research Online. Washington, D.C. October 27, 2017.

Smiraglia, R. P., & Ridenour, L. (eds.). (2017). Proceedings from the North American Symposium on Knowledge Organization. Vol. 6. Champaign, IL. http://www.iskocus.org/nasko2017-proceedings.php.

BOOK CHAPTER

Ginda, M. P., Gniady, T., Boyles. M. J., Ridenour, L. E. (2014). Using Point of View Cameras to Study Student-Teacher Interactions. In K. Börner, D. E. Polley (Eds.), Visual Insights: A Practical Guide to Making Sense of Data pp. 248-253. Cambridge, MA.

INVITED BOOK REVIEWS

Ridenour, L. (2020, forthcoming). Organizing Library Collections: Theory and Practice [book review]. Journal of Education for Library and Information Science.

Ridenour, L. (2016). Indexing It All: The Subject in The Age of Documentation, Infor-

mation, And Data [book review]. Knowledge Organization 43, 306-309.

OTHER PRINT CONTRIBUTIONS

Smiraglia, R. (2015). Domain Analysis for Knowledge Organization: Tools for Ontology Extraction. Netherlands: Elsevier Science. [visualizations for chapters 2 and 6]

PRESENTATIONS

Ridenour, L. (2020). Origami of Knowledge: Folding Data to Identify Shared Concepts. University of Wisconsin-Milwaukee 3 Minute Thesis Preliminary Round. Milwaukee, WI.

Ridenour, L. (2020). Epistemic Origami: Using Data Science to Illuminate Concepts. School of Information Studies PhD Workshop: Pitching Our Research, Milwaukee, WI.

Ridenour, L. (2019). Examining the Notion of the Boundary Object in Information Systems: The Transdisciplinary Oeuvre of Cognitive Science [Doctoral Poster Competition]. Association for Library and Information Science Education 2019 Annual Conference, Knoxville, TN.

Ridenour, L. (2019). Examining the Notion of the Boundary Object in Information Systems: The Transdisciplinary Oeuvre of Cognitive Science [Doctoral Colloquium]. North American Symposium on Knowledge Organization. Philadelphia, PA.

Ridenour, L. (2019). Examining Concepts Across Boundaries: A Pre-Processing Framework. North American Symposium on Knowledge Organization, Philadelphia, PA.

Ridenour, L. (2019). Updating Medieval Manuscript Data Using Supervised Classification: A Preliminary Process. iConference. Washington, D.C.

Sainte-Marie, M, & Ridenour, L. (2018). Examining Communities in the Transdisciplinary Area of Cognitive Science: Automatic Classification for Examining Communi-

ties in the Web of Science Using Unsupervised Clustering Methods. Presented at the 29th ASIS&T Special Interest Group/Classification Research (SIG/CR) Workshop on Culture, Community, and Voice in Knowledge Organization Systems, Vancouver, B.C.

Rankowski, B., & Ridenour, L. (2018). Navigating Medical Relations Along the Autism Spectrum. Presented at the Self-Advocates Preconference at the 50th Autism Society of America Annual Conference. Bethesda, MD.

Ridenour, L. (2017). Visualizing Data, Information, and Knowledge: A Discussion. North American Symposium on Knowledge Organization, Champaign, IL.

Ridenour, L., & Kozak, N. (2018). Framing Shifts in Domains and Terminology as Concept Repurposing: A Theoretical Lens. Presented at the School of Information Studies Research Brown Bag Series, Milwaukee, WI.

Freeburg, D., Gamble, A., Ridenour, L. (2017). The ASIS&T Living Manual: Understanding SIG Participation. Presented at the ASIST 2017 SIG Cabinet Meeting, Washington, D.C.

Ridenour, L., & Zolyomi, A. (2017). Breakout Session: Beyond Classifying Neurodiversity. Presented at the 28th ASIS&T Special Interest Group/Classification Research (SIG/CR) Workshop on Building a Research Agenda for Multi-Perspective Knowledge Representation, Washington, D.C.

Ridenour, L. (2017). Boundary Objects and Interfield Theory: Examining these Notions in Information Systems. Presented at the School of Information Studies Research Brown Bag Series, Milwaukee, WI.

Ridenour, L. (2016). Leveraging the Power of Social Reading and Big Data: An Analysis of Co-Read Clusters of Books on Goodreads. iConference, Philadelphia, PA.

Ridenour, L. (2016). Visualization of Co-Read Book Data at Goodreads.Com: A Potential Readers' Advisory Tool. Association for Library and Information Science Educa-

tion 2016 Annual Conference, Boston, MA.

Ridenour L. (2015). Case Study of Waiting List on WPLC Digital Library [Poster Session, Minute Madness Participant]. Joint Conference on Digital Libraries, Knoxville, TN.

Ridenour, L. (2015). Concept Term Repurposing: Framing Shifts in Domains and Terminology. North American Symposium on Knowledge Organization, Los Angeles, CA.

Ridenour, L. & Park, H. (2014). Visual Analysis of Historical Patterns: ISKO Proceedings Titles from 1992 to 2012. International Society for Knowledge Organization. Kraków, Poland.

INVITED TALKS

Ridenour, L. (2020). Building a Working Computer for under $100 [Community Engagement]. Milwaukee County Federated Library System Adult Reference Meeting: Tech Show and Tell. Milwaukee, WI.

Ridenour, L. (2020). Updating Medieval Manuscript Data Using Supervised Classification: A Discussion. LEADS Data Science Symposium, Philadelphia, PA.

Ridenour, L. (2016). Boundary Objects as Interfield Phenomena: From Sociological Phenomena To Information System Artifacts. COST KNOWeSCAPE Workshop on Alternative and Tailored Metrics. Warsaw, Poland.

PREVIOUS PROFESSIONAL EXPERIENCE

RESEARCH EXPERIENCE

LEADS-4-NDP Data Science Fellow 2018

Drexel University, University of Pennsylvania

Dot Porter and Doug Emery, supervisors

Research Assistant 2018

University of Wisconsin-Milwaukee

Professor Richard Smiraglia, supervisor

Research Assistant 2014 – 2015

University of Wisconsin-Milwaukee

Professor Richard Smiraglia, supervisor

Research Assistant 2013 – 2016

University of Wisconsin-Milwaukee

Professor Wooseob Jeong, supervisor

Research Assistant 2013

Indiana University

Professor Katy Börner, supervisor

TEACHING EXPERIENCE

Lecturer (instructor of record, onsite, two sections) Fall 2019

INFOST315: Knowledge Organization for Information Science and Technology

Lecturer (instructor of record, online compressed course) Summer 2019

INFOST315: Knowledge Organization for Information Science and Technology

Lecturer (instructor of record, onsite) Spring 2019

INFOST315: Knowledge Organization for Information Science and Technology

Lecturer (instructor of record, onsite) Fall 2018

INFOST230: Organization of Knowledge

Teaching Assistant Spring 2017

INFOST511: Organization of Information (online)

Professor Richard Smiraglia, supervisor

Teaching Assistant Fall 2016

INFOST110: Introduction to Information Studies

Professor Nadine Kozak, supervisor

STUDENT MENTORING

"Examining Knowledge Organization Systems: The Pokédex" 2019

Jacob Surch, Mentee

GUEST LECTURES

Doctoral Seminar on Knowledge Organization Spring 2018

Topic: Domain Analysis

Information Studies 990: Mixed Methods Doctoral Seminar Fall 2017

University of Wisconsin-Milwaukee

Topic: Mixed Methods Research Design

Introduction to Information Studies Fall 2017

University of Wisconsin-Milwaukee

Topic: Networks (information and computer); Web 2.0

Knowledge Organization Doctoral Seminar Spring 2016

University of Wisconsin-Milwaukee

Topic: Knowledge Organization and Interdisciplinarity

FUNDING AND AWARDS

COVID-19 Academic Support Grant 2020

Academic Mamas Foundation

$500

Jean Tague-Sutcliffe Doctoral Student Research Poster Competition: 2019

Third Place (tie)

Association for Library and Information Science Education

Eugene B. Garfield Doctoral Dissertation Fellow 2019

Beta Phi Mu Honor Society

$3,000

LEADS-4-NDP Data Science Fellow 2018

Drexel College of Informatics and Computing, Metadata Research Center

Site: University of Pennsylvania Libraries, mentors: Dot Porter and Doug Emery

$5,000 stipend, $1,000 conference travel funding

Chancellor's Fellowship 2017 – 2018

University of Wisconsin-Milwaukee

$1,000 scholarship

Advanced Opportunity Program Fellow 2016 – 2019

University of Wisconsin-Milwaukee

$15,000 yearly stipend, $1,000 conference travel funding

Chancellor's Fellowship 2017

University of Wisconsin-Milwaukee

$1,000 scholarship

New Leaders Award 2016 – 2018

Association for Information Science and Technology

Assignment: SIG/CR, mentor: Barbara Kwaśnik

$1,000 conference travel funding ($700, $300), waived conference fees

Chancellor's Fellowship 2013 – 2014

University of Wisconsin-Milwaukee

$1,500 scholarship

Homer E. Marsh Scholarship 2012 – 2013

Indiana University Bloomington

$1,000 scholarship

FUNDED COLLOQUIA

LEADS-4-NDP Forum 2020

LIS Education and Data Science for the National Digital Platform

$1,000

Doctoral Colloquium 2019

North American Symposium on Knowledge Organization

$200

Doctoral Colloquium 2019

iConference

$1,300 travel funding, waived conference fees

Doctoral Colloquium 2018

Association for Information Science and Technology

Waived conference fees

Doctoral Colloquium 2017

North American Symposium on Knowledge Organization

$250

PROFESSIONAL AFFILIATIONS

Member 2014 – 2016, 2020 – present

American Library Association

Member 2014 – 2016, 2020 – present

Wisconsin Library Association

Member 2013 – present

Knowledge Organization Research Group (KOrg)

University of Wisconsin-Milwaukee

Member 2014 – present

International Society for Knowledge Organization

Member 2012, 2015 – present

American Society for Information Science & Technology

SERVICE

WORKSHOPS AND CONFERENCES

Dobreski, B., & Ridenour, L. Knowledge Organization: Community and Computation. Conference. North American Symposium on Knowledge Organization. Philadelphia, PA. June 13-14, 2019.

Montoya, R. D., Hajibayova, L., Matsuda, S., Ridenour, L. Zhitomirsky-Geffet, M. Culture, Community, and Voice in Knowledge Organization Systems. Workshop. 81st Annual Meeting of the Association for Information Science & Technology. Vancouver, B.C. November 10, 2018.

Smiraglia, R. P., & Ridenour, L. Visualizing Knowledge Organization: Bringing Focus to Abstract Realities. Conference. North American Symposium on Knowledge Organization. Champaign, IL. June 15-16, 2017.

Hajibayova, L., Kwaśnik, B. H., Montoya, R. D., Ridenour, L., and Zhitomirsky-Geffet, M. Building a Research Agenda for Multi-Perspective Knowledge Representation. Workshop. 80th Annual Meeting of the Association for Information Science & Technology. Washington, D.C. October 27, 2017.

REVIEWING AND OTHER PUBLICATION ACTIVITY

Reviewer 2020

Association for Information Science and Technology Conference

Reviewer 2020

Knowledge Organization

Reviewer 2019

Knowledge Organization

Reviewer 2019

iConference

Reviewer 2019

Association for Information Science and Technology Conference

Reviewer 2018

Knowledge Organization

Editor 2017

SOIS Ph.D. Newsletter

School of Information Studies, University of Wisconsin-Milwaukee

Editor 2015

SOIS Ph.D. Newsletter

School of Information Studies, University of Wisconsin-Milwaukee

Editor in Chief 2014

SOIS Ph.D. Newsletter

School of Information Studies, University of Wisconsin-Milwaukee

Editor 2013

SOIS Ph.D. Newsletter

School of Information Studies, University of Wisconsin-Milwaukee

LEADERSHIP

Chair Elect 2019 – 2020

SIG/CR, Association for Information Science and Technology

President 2017 – 2019

International Society for Knowledge Organization, Canada/United States

Secretary/Treasurer 2016 – 2019

SIG/CR, Association for Information Science and Technology

Social Media Officer 2016

SIG/CR, Association for Information Science and Technology

President 2016

Doctoral Student Organization, School of Information Studies

University of Wisconsin-Milwaukee

Communications Officer 2015 – 2016

SIG/CR, Association for Information Science and Technology

Student Representative 2015 – 2016

Doctoral Programming Committee

School of Information Studies, University of Wisconsin-Milwaukee

President 2012 – 2013

American Society for Information Science and Technology – Student Chapter

Indiana University, Bloomington, Indiana

VOLUNTEER WORK

UW-Milwaukee Pavilion Volunteer 2015

American Library Association Annual Conference

UW-Milwaukee Pavilion Volunteer 2014

American Library Association Annual Conference

CONSULTING WORK

i3 Inclusion Institute 2019

Topic: Diversity Recruitment

ADDITIONAL EDUCATION

CERTIFICATION COURSEWORK

Online and Blended Teaching Program 2020

Center for Excellence in Teaching and Learning

University of Wisconsin-Milwaukee