2017

Seeking social capital and expertise in a newly-formed research community: a co-author analysis

Christine E. Forte

Pepperdine University

Graduate School of Education and Psychology

SEEKING SOCIAL CAPITAL AND EXPERTISE IN A NEWLY-FORMED RESEARCH COMMUNITY:

A CO-AUTHOR ANALYSIS

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Education in Learning Technologies

by

Christine E. Forte

November, 2017

Judith Fusco Kledzik, Pd.D. -- Dissertation Chairperson

This dissertation, written by

Christine E. Forte

under the guidance of a Faculty Committee and approved by its members, has been submitted to and accepted by the Graduate Faculty in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

Doctoral Committee:

Judith Fusco Kledzik, Ph.D., Chairperson

Linda Polin, Ph.D.

Timothy Podkul, Ph.D.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Many thanks to my DELT-17 classmates for their great thinking, commiseration, support, and laughter; somehow we managed fit a full-time graduate program into our already full lives. Thanks to the many excellent and caring instructors who challenged us with interesting readings and creative assignments; were this work not so brutal, I would not have wanted it to end. Thanks also, to my employer, Antioch University; the acceptance of my unconventional sabbatical proposal made the required travel in this program possible.

I'm grateful to my committee members, Judi Fusco, Linda Polin, and Tim Podkuhl, who gave generously of their time and expertise. Their patience with multiple drafts and the questions they asked immeasurably improved this document and deepened my understanding.

Dissertation work is difficult and lonely; encouragement from friends, colleagues, and family made the journey more bearable. I owe a special debt of gratitude to Dawn Murray for her empathy, unfailing emotional support, and spot-on advice during our beach walks. Attending graduate school and completing my dissertation would not have been possible without Eric Forte's excellent, flexible co-parenting, and good humor. Vaughn and Zane Forte grew into young men while I went to school and wrote this dissertation; they were and remain sources of inspiration and joy.

VITA

EDUCATION

2017	Doctor of Education in Learning Technologies	Pepperdine University Graduate School of Education and Psychology Malibu, CA
1994	Master of Science Library and Information Science	University of Illinois at Urbana Champaign School of Information Sciences Champaign, IL
1991	Bachelor of Arts Germanic Languages and Literatures	University of Illinois at Urbana Champaign Champaign, IL

PROFESSIONAL EXPERIENCE

10/2005 – Present	Director of Library Services; Core Faculty Antioch University Santa Barbara Santa Barbara, CA
8/2000 -- 8/2007	Reference Librarian Santa Barbara Public Library Santa Barbara, CA

ABSTRACT

This exploratory study applies social network analysis techniques to existing, publicly available data to understand collaboration patterns within the co-author network of a federally-funded, interdisciplinary research program. The central questions asked: What underlying social capital structures can be determined about a group of researchers from bibliometric data and other publicly available existing data? What are ways social network tools characterize the interdisciplinarity or cross-disciplinarity of co-author teams? The names of 411 grantees were searched in the Web of Science indexing database; author information from the WoS search results resulted in a 191-member co-author network. Research domains were included as attribute data for the co-author network. UCINet social network analysis software calculated a large 60 node component and two larger components with 12 and 8 nodes respectively, the remainder of the network consisted of smaller 2-5 node components. Within the 191-node co-author network the following analyses were performed to learn more about the structural social capital of this group: Degree and Eigenvector centrality measures, brokerage measures, and constraint measures. Additionally, ten randomly selected dyads and the five 4-node cliques within the 191-node network were examined to find patterns of cross-disciplinary collaboration among researcher and within award teams. Award numbers were added as attribute data to five 4-node cliques and 10 random dyads; these showed instances of collaboration among interdisciplinary award teams. Collaboration patterns across disciplines are discussed. Data from this research could serve as a baseline measure for growth in future analyses of the case studied. This method is recommended as a tool to gain insights to a research community and to track publication collaboration growth over time. This research method shows potential as a

way to identify aspects of a research community's social structural capital, particularly within an interdisciplinary network to highlight where researchers are working well together or to learn where there is little collaboration.

Chapter 1: Introduction

Interdisciplinary research provides a deeper understanding of scientific knowledge and enables scientists to tackle problems that lie beyond the scope of a single discipline. As a body of scientific research matures, there is an increase in scientific knowledge and technological advancements enabling exploration into ever more sophisticated questions. Interdisciplinary research is more than a means to advance science for the sake of advancement. Interdisciplinary research and multi-disciplinary research were born out of a need to bring multiple disciplines and research methods to bear on complex scientific and social problems (Committee on IDR, 2004; Ledford, 2015). Many problems facing our world today (such as climate change, energy and food production, global health, and population displacement caused by political or environmental reasons) are so complex and multifaceted that they demand interdisciplinary approaches, if solutions are to be found This evolution of science, coupled with corresponding world and social complexity presents researchers with problems that can no longer be answered within the framework of a single discipline (Committee on IDR, 2004; Ledford, 2015; Wagner, 2008). Award granting bodies interested in finding solutions to complex problems encourage interdisciplinary research by soliciting proposals from interdisciplinary teams and establishing interdisciplinary centers.

Over time, curious, passionate scientists have made forays into other disciplines to solve problems. (e.g., Pierre and Marie Curie) and groups of researchers sometimes begin to work together of their own accord (Wagner, 2008). In most modern cases though, some group, body, or office nurtures interdisciplinary relationships through funding resources. Interdisciplinary research labs or centers may be funded by Universities, philanthropic organizations with a research agenda, or government agencies. Two examples of funding organizations with a wellness mission are the Robert Wood Johnson Foundation and the John A. Hartford Foundation. The Robert Wood Johnson Foundation's Interdisciplinary Research Leader's program funds three person teams (two researchers and a community member) whose research project proposes to tackle complex local problems and to create a "culture of health" (n.d.) within the community (Interdisciplinary Research Leaders). Keeping people healthy as they age is a complex problem; the John A. Hartford Foundation began supporting interdisciplinary research in health care in 1990 and in 2004, provided matching funds to create 12 Geriatric Care Research Centers across the United States, with a focus on interdisciplinary research and interdisciplinary leadership (Wetle & Pincus, n.d.). Other organizations, like the MacArthur foundation, support multiple interdisciplinary research networks out of an interest in addressing current social issues and providing empirical evidence that can serve as a basis for changes in policies and practices (MacArthur Foundation, 2017). Federal government organizations like the Department of Defense, National Institutes of Health, and the National Science Foundation periodically create funding programs for interdisciplinary research. Successful interdisciplinary centers within universities may arise from faculty who see a need for interdisciplinary collaboration, from administrators who envision interdisciplinary research as a part of a strategic plan, or a hybrid of the two (Association of American Universities, 2005; Strober, 2015). A funding agency may also sponsor centers, institutes, or consortia that support a community of interdisciplinary researchers who are also stakeholders in the outcomes of a project (Committee on IDR, 2004). (e.g., The Institute for Mathematics and Its Applications at the University of Minnesota and The Beckman Institute at the University of Illinois, Urbana

Champaign) Stakeholders in interdisciplinary and multi-disciplinary research, especially stakeholders responsible for publicly funded projects, want to make their research community's membership, expertise, and knowledge transparent. These stakeholders have a responsibility to learn whether the anticipated outcomes of funded research programs are on-track, if goals have been achieved, and if funding is reaching the right people. One of the challenges facing interdisciplinary funding agencies in new areas of research is building community among the funded researchers. Communication across disciplines is hard work, cross-disciplinary communication and interdisciplinary collaboration requires an understanding of the research culture, methods, and vocabulary of another domain, which are not easy barriers to overcome (Strober, 2015). This research project seeks to analyze and make visible the co-authorship ties among grantees from a newly funded National Science Foundation program as a way of showing some measurable communication and collaboration among those grantees. Co-authorship was selected, because the publication of a research article can be seen as a successful outcome, article publications serves as artifacts of successful researcher collaborations, and research articles can be found using publicly available resources.

NSF created the Cyberlearning: Transforming Education (C:TE) program from 2011-2014 and continues to shape the meaning of Cyberlearning (CL)through the projects it funds. In 2014, the CL program area was renamed to Cyberlearning and Future Learning Technologies (C:FLT). Multiple directorates across NSF-funded C:TE and currently fund C:FLT. NSF defines cross-cutting as those "Activities in which two or more NSF directorates and/or other federal agencies participate" (n.d., para. 7). The program solicitations for both C:TE and C:FLT stated the expectation that all research teams would contain interdisciplinary expertise (NSF, 2011;

2014). Four different award types were available through C:TE and C:FLT to fund projects of various scope and length; a fifth category was a single award to create a Center to support all CL projects (NSF, 2011). The cross-cutting nature of Cyberlearning, coupled with the newness of this research create some challenges in making the community transparent.

For example, NSF – a major United States funding agency – is divided up into divisions which fund various programs. Program Officers exist for every program. NSF-funded “centers are a principal means by which NSF fosters interdisciplinary research” (NSF, 2011, p. 389). These groups are the stakeholders with a vested interest in the outcomes of a program. (In cases of privately funded research, stakeholders are composed of different groups.) One problem facing stakeholders of interdisciplinary research is the difficulty in gauging the growth and diffusion of interdisciplinary, cross-cutting research.

In the case of C:TE and C:FLT (hereafter referred to simply as CL as a broad term for NSF Cyberlearning programs, awards, or grantees), the Center for Innovative Research in Cyberlearning (CIRCL) was founded to “support, synergize, and amplify” (para.1) the efforts of CL researchers (CIRCL, n.d.-a). The CL program officer and CIRCL, therefore, are major stakeholders in CL research. The C:TE grantees were the inaugural group of NSF CL researchers. Although the CL program is relatively young at the time of this research, it is not too early to begin to map the expertise in this research community and to begin to outline the relationships between these researchers

In much the same way that Rogers (2003) documented ideas moving through and being adopted by members of a community, the evolution of science follows a logistic curve (Ben-David & Collins, 1966). During early years of a newly funded area of research, when scholars are just coming to the field, many of the informal structures through which knowledge can flow or be shared among a group have not been built. For example, organizations might ordinarily report demographic information about a community or host discussion forums (which could be mined for data), but in newly funded areas, these may not yet exist. Even if such virtual spaces were dedicated to a new area of research, the researchers themselves may not yet have the cohesiveness as a community or a sense of common purpose to want to communicate with one another there. Multi-year projects may not yet have data to report, and those pioneer researchers who do have data to report may not find specialized journals dedicated to their topics. For these reasons, a new research community may remain somewhat invisible, in the early years. A multi-disciplinary field may struggle more, since conversations and research could take place in a variety of domains. Stakeholders in interdisciplinary research may find it difficult to know the scientists in the community, in which domains they research, how they are working together, and to understand how these researchers might be collaborating beyond the boundaries of their awards.

Research Purpose

The CL research community was selected as a case of a newly formed interdisciplinary research community. In such emerging research areas, few artifacts are publicly available for study. In the case of CL research publicly available documents include: NSF awards granted, papers written, conferences attended, and data collected by and shared by the Center for Innovative Research in Cyberlearning (CIRCL) on their website; though the sources of data may vary, these types of information are publicly available for many funded areas of research. This project lays a foundation for future work by studying the connections of the members of the funded CL community through their co-author collaborations; this type of data can provide a point of reference for growth, impact, and diffusion in following years. This study also provides a proof of concept of what social structural information be learned about a research community through publicly available data.

Social network analysis tools were applied to bibliometric data to calculate the social structural capital within this community; these calculations were supported by graphs that visualize the relationships between CL grantees. UCINet (Borgatti, Everett & Freeman., 2002) social network analysis software calculated relationships between members of a community to surface a variety of network phenomena. Graphs created with visualization software incorporate colors, shapes, hues, varying node sizes, and varying line thicknesses to represent qualitative data and aid the reader's understanding of the network analyses. In this way, social network analysis software can provide insight into the structural social capital of a newly forming research area by studying the co-author network of a group of scholars; in this case, the co-author relationships between CL PI/Co-PIs from the initial round of NSF CL funding (2009-2014).

Theoretical Framework

When stakeholders ask about the connections among researchers and how the research community is growing, they are inquiring about structural social capital. The theory of social capital, in its simplest terms, can be explained by the adage: It's not what you know, it's who you know. This maxim emphasizes the importance of social connections and the way relationships can be assets to individuals. Social capital is the potential or real value (i.e., capital) that can be realized because of human relationships (Bourdieu, 1983; Coleman, 1994; Putnam, 2001). Lin (2001) states, "social capital consists of the resources embedded in relations and social structure" (p. 24). The structure of social capital (i.e., who knows whom) is an indicator of how ideas and knowledge spread throughout a group (Burt, 1992).

Social network analysis tools analyze and quantify the social structural capital with the parameters of a network. These calculations can reveal the individuals who are positioned at the center of a network, if clusters of individuals are working together, and which individuals are poised to bridge a gap between unconnected groups. These relationships are most easily seen via network graphs, where individuals are typically designated as vertices (or nodes) and individuals with a relationship are typically connected by an edge (a tie). These graphs are called sociograms in social network analysis. In a sociogram, relationships between individuals become visible and clusters (or components) of connected individuals may emerge. These clusters and the connections between them can represent the ways in which information flows throughout a community. Burt (1992) refers to unconnected or loosely connected components as *structural holes*. Structural holes are an important concept in Social Capital. Sometimes, sociograms (created with social network analysis tools) make visible those individuals who have connections with members outside their own clusters. Those, who bridge across structural holes, are potentially key members in the movement of information and ideas across a community (Burt, 1992). Stakeholders who are interested in how information is flowing through a research community would have an interest in learning the members who could potentially unite various groups.

Seeing co-author relationships among members of a research community is potentially useful for stakeholders, because it provides data about which researchers are working with whom – beyond the collaboration of an award or grant or other funding. By including attribute data about the domain where researchers work, co-author relationships provide insight to the interdisciplinarity of collaborative writing teams, another data point of interest to stakeholders in interdisciplinary funding. Graphs of researcher relationships can provide insight to stakeholders about the researchers who are potentially key to spreading new ideas (because they bridge structural holes) and those who are nurturing existing ideas among teams of researchers (because of a central location within a network). In this study, social network analysis tools were used to show the researchers and the domains that are particularly well-connected in the (co-author) community of CL research, as well as the researchers (and what domains they represent) who occupy structural holes and other potential brokering positions within the network. Such information could be useful to stakeholders who wish to understand the collaborations of researchers from various domains in an interdisciplinary community. Such information could also be used by stakeholders to shape the flow of information in a community by fostering connections among researchers across those domains where stakeholders would like to see more collaboration.

Methods

Anyone with a desire to learn about the connections between researchers within an interdisciplinary community and how that research community is growing is asking questions about the structural social capital of that community. Social network analysis is a method to make the social capital within a group visible. Social network analysis makes structural relationships apparent and allows stakeholders "to measure and represent these structural relations accurately, and to explain both why they occur and what are their consequences" (Knoke & Yang, 2008, Section 2.1, para. 2). A true social network analysis requires considerable time and participation from the community. Research scientists, especially those who win NSF awards, are likely to be high-level researchers and deeply involved with their careers. Conducting a true social network analysis of the CL community may be a burden to these researchers. It is possible, though, to employ social network analysis techniques to existing data and make some relationships within a research community visible, without directly contacting members. Indeed, with the advent of Web 2.0, there are many existing artifacts that can act as a proxy for direct contact with researches. Some examples of the kinds of publicly available data to which social network analysis techniques can be applied are: blogs, micro-blogs, listserv conversations, and open source software development. This study used bibliometric data, specifically co-author relationships, to create a snapshot of relationships in the early stages of CL research.

Social network analysis techniques are commonly applied to co-author relationships, because co-authorships answer the question: With whom? (Zervas, 2014). There are some weaknesses in using social network analysis techniques in this way; the primary problems are that it is impossible to quantify the relationships among authors (Jeanes, Loacker, & Śliwa, 2014). Co-authorship is also an imperfect measure of writing collaboration, because sometimes co-authorships are given to those who secured funding or played a minor role in writing the paper (Katz & Martin, 1997; Laudel, 2002). With these limitations in mind, co-authorship analysis continues to be used as a measurement of collaboration, barring other more accurate ways of measuring collaboration (Harmelen, 2012). Numerous examples of co-author data being used in combination with social network analysis techniques exist in the literature. (See chapter two for more details.)

Significance

This research is designed to lay the foundation for future research on the growth and the constitution of interdisciplinary research domains based on co-author relationships. Researcher domain and the variety of domains represented within co-authors teams can be shown in network graphs when researchers' domains are included as attribute data for the network. Data collected in this study could be used as a comparison in a second, future co-authorship analysis that reexamines the social structural capital of co-author relationships within this research community. Stakeholders looking for indicators of multidisciplinary collaboration may find a co-authorship analysis useful, since "co-authorship is a prominent indicator of collaboration in scholarship" (Rodriguez & Pepe, 2008, p195). Co-authorship analyses have been used to track the growth from the early stages of various research fields (Börner, Dall'Asta, Ke, & Vespignani, 2005; Guan & Liu, 2014; Yin, Kretschmer, Hanneman & Liu, 2006), this study adds to that body of work.

Second, should the results of this study be valuable to stakeholders, the use of publicly available tools and data make this study replicable. The use of publicly available research artifacts allows for the data collection and analysis to potentially become (more) automated. The design of this study also sheds light on the collaborative relationships (in the form of publications) between scientific researchers in an unobtrusive, non-invasive manner; this is important when the relationships being studied are those of high-level research scientists. Potentially, stakeholders could employ this type of research into formative and summative evaluations of the growth of a funded research area. An analysis of co-author relationships may be of value to long-term funding, such as research centers or interdisciplinary departments within a university.

Research Questions

This study investigated two research questions:

- RQ1: What underlying social capital structures can be determined about a group of researchers from bibliometric data and other publicly available existing data?
- RQ2: What are ways social network tools characterize the interdisciplinarity or cross-disciplinarity of co-author teams?

Research Design

This study explored the structures of social capital in a research community by applying social network analysis techniques to the co-authorship network of grant recipients from an NSF funded program. A true social network analysis requires considerable time and participation from the community. Research scientists, especially those who win NSF awards, are likely to be high-level researchers and deeply involved with their careers. Therefore, conducting a true social network analysis would impose a burden upon these researchers and would not likely yield good results. It is possible to implement social network analysis techniques on existing data to make some relationships within a research community visible without directly contact members, hence the purposeful decision to use existing publicly available data.

This research took place in three stages: Identifying CL researchers and collecting qualitative data used to establish the research domains of members of this community;

collecting publication citation data of the CL researchers; establishing the co-author collaboration network of CL researchers and adding attribute data based on the qualitative data from stage one; analyzing the data using basic social network analysis measures to calculate centrality measures and identify those nodes in structurally significant brokering positions within the network; creating graphs to visualize the data that was analyzed.

Data for this research was obtained from an NSF-funded, publicly available data mining tool that allows searches NSF researchers, projects and awards: Deep Insights Anytime Anywhere (DIA2, n.d.). Qualitative information used to establish researchers' domains, employment affiliations, final academic degrees and links to CV's (where possible) was collected from the open Web. The second stage of research employed Google Scholar (a scholarly search engine) and Thomson Reuters' ciation index Web of Science Core Collection (WoS). The names of Principal Investigators and Co-Principal Investigators (PI/Co-PIs) from CL were searched within WoS and CL award numbers were searched in Google Scholar. Bibliometric data from these search results were downloaded and co-author collaborations among the CL researchers formed the CL researcher network. During the first and second steps of the research, the target population was bounded by the names of CL researchers. In the final step of this research, the population was bounded by those CL researchers who were found to have co-authored with other CL researchers between the years 2009-2015.

UCINet a social network analysis software, was used to calculate the structural capital and UCINet's visualization tool NetDraw were combined to create graphs that illustrate the structure of the CL researchers who published with other CL researchers (Borgatti et al., 2002). Departmental, academic degree and information about research interests (where available) collected from websites provided the attribute data that allowed cross-domain co-authorship collaborations to be seen.

Assumptions, Delimitations, Limitations, and Scope

A key assumption in this research is that scientific knowledge is socially constructed (Crane, 1972; McFadyen & Cannella, 2004; Pinch & Bijker, 1984). Through collaborations and conversations, researchers build an understanding of their problem space, research questions, and methodology. If researchers understand, build, and grow their respective fields together then it makes sense to look at relationships among scientists in terms of social capital. There are a variety of ways in which scientific relationships can be made visible.

Research teams and collaborative writing teams are indicative of relationships between researchers. It is not possible to conduct research with someone or write a paper with them without developing a relationship. Repeated co-authorships or research projects are evidence of stronger bonds, since researchers voluntarily enter into these agreements. One assumption is that collaborative writing and research team membership reflects a certain level of trust. That these relationships are also a vehicle through which information and thinking about developments in CL research flows through the community of scientists, who are exploring this nascent area of research is another assumption. Finally, this research assumes that published papers and PI/Co-PI team membership are artifacts of these relationships and can therefore be used as a proxy to make the social structural capital of this group of researchers visible.

Several choices were made that to narrow the focus of this research. First, co-author relationships among CL researchers were used to illustrate the structural social capital in this community. Co-authorships can be used to illustrate relationships and relationship patterns in a research community and these databases are well-documented (Newman, 2004). Second, the publications studied were limited to those found in Thomson Reuters' Web of Science (WoS) Core Collection; this database provides interdisciplinary coverage by combining the Social Science Citation Index, the Science Citation Index, the Conference Proceedings Citation Index, along with several others (Thomson Reuters, 2014).

Thomson Reuters' WoS indexes leading scholarly journals (Thomson Reuters, 2014). Relying primarily on WoS for evidence of co-author collaborations limits the scope of data collected to top-tier academic journals; this fails to accurately represent the cycle of academic scholarly communications. White papers, technical papers, position papers, and other prepublications may serve as the opening salvos to an academic conversation or to inform a research community of one's work (Library of Congress, 2008); these writings often lay groundwork for articles that may eventually be published in competitive academic journals. These early writing collaborations can also be the first steps in research that lead to grant funding and further collaborations (Technology Innovation Program, 2010). This grey literature could serve as an early indicator of community formation, but are not captured by the data sources used in this study. Despite this limitation, the accuracy and consistency of WoS made the index a natural fit for this research; a project of this size does not have the resources needed for data clean up were other sources to be used. Although the data from WoS cannot provide as complete a picture of the CL community's co-author relationships, the WoS data serves as a proof of concept for this data collection method.

Not only are some publications omitted from this analysis because they did not appear in WoS, the co-authorship relationship itself offers a limited view of collaboration. Not all authors are necessarily writers, and not all of those who supported the research behind the publications are listed as authors (Laudel, 2002). Additionally, there may be other ways that researchers collaborate that do not leave behind evidence as easily and commonly uncovered as co-authorships (e.g., casual conversations at a conference, use of a lab).

As part of the second phase of this research, the names of all CL PI/Co-PIs were searched in the WoS database, and bibliographic data from all their publications from 2009-2015 were downloaded; the date range 2009-2015 is also the time during which CL awards were distributed. There are two main limitations associated with this choice. First, it is so early in the life of the CL project, that there are projects without findings to release. Second, it may also occur that some of the articles downloaded from 2009 were written by researchers who did not receive an award until 2015; this certainly shows a co-author collaboration, and a relationship but it is not a collaboration stemming from the award itself, which is an implicit assumption of this research.

The research took place at the meso level, encompassing between 101-10,000 records (Börner & Polley, 2014) with 411 CL researchers. The study included network analyses (co-author collaborations), topical analysis (research domain), and descriptive statistics about the researchers. The study used applied social network analysis techniques to reveal the structural social capital inherent in the CL community, as revealed through co-authorship and research team membership. Terms used in this research are defined in the following section.

Definition of Terms

This section contains some of the key terms used in this study:

Actor: The entities that comprise a network system; these can be individuals or collectives (Borgatti, Everett & Johnson, 2013).

Adjacency: A term in graph theory to describe a direct connection between two Nodes (Scott, 2013)

Adjacency Matrix: The mathematical representation of the direct relationships between actors in a network (Borgatti et al., 2013; Scott, 2013). Data from adjacency matrices is used to calculate social analysis measures such as centrality and homophily.

Alter: Any node connected to the node that is the focus of analysis or discussion (Borgatti et al., 2013).

Attributes: Characteristics of Actors or Nodes; attributes can be categorical or quantitative (Borgatti et al., 2013).

Betweenness Centrality: A measure of how often a node is found on the shortest path between any other two nodes in a network. This analysis based on position can reveal which nodes are gatekeepers in a community; a high betweenness number for Node X indicates that many nodes rely Node X to communicate with others in the network (Borgatti et al., 2013).

Broker: A node that connects two otherwise unconnected nodes in a network can be said to occupy the position of broker (Fleming, Mingo, & Chen, 2007). Using Burt's (1992) theory of structural holes, a broker spans a structural hole in a network.

Centrality: A group of formulas that measure an individual Nodes contribution to a network based on a nodes position within the network; the centrality of any given node provides an indication of potential opportunities available to a node because of its structural position within the network (Borgatti et al., 2013).

Clique: A network subgroup made up of nodes that are connected with one another; this number must be at least three (Borgatti et al., 2013).

Closeness Centrality: A measurement based on the structure of a network that provides an indication of how easily information can flow through a network; this measurement is not meaningful in fragmented networks (Borgatti et al., 2013).

Component: "A maximally connected sub-graph" (Scott, 2013, p. 100). A component is a group of nodes in which it is possible to trace a path from each member to all others in the component. Finding components is one of the first steps in analyzing the structure of a network (Scott, 2013). A component is an indicator of social capital within its nodes. At the early stages of network formation, components can also be early indicators of collaboration.

Degree: The number of other Nodes to which Node X is connected in a graph (Borgatti et al., 2013; Scott, 2013).

Degree Centrality: A measurement of the connectedness of individual nodes in a network graph; this measurement is relevant to the size of a graph and based on a node's position within the network. Essentially the number of other nodes to which an ego is connected (Borgatti et al., 2013; Scott, 2013).

Density: A measurement of the relative number of ties in a network (Scott, 2013).

Directed Network Graph: A graph in which the relationships between nodes need not be reciprocal (e.g., trust or friendship) (Borgatti et al., 2013; Scott, 2013).

Dyad: The relationship between two actors in a network; dyads are the "fundamental unit of data collection" (Borgatti et al., 2013, p. 2).

Edge: See Tie

Ego: The node that is the focus of an analysis. Other nodes connected to the Ego are called Alters (Borgatti et al., 2013).

E-I Index (External-Internal Index): A measurement of homophily among groups that calculates the number of ties among groups (internal) and the number of ties between groups (external). E-I Index scores range between -1 (internal groups maximally connected) and 1 (external groups maximally connected) (Hanneman & Riddle, 2005).

Homophily: The tendency for those with similar characteristics to find affinity and connection with one another (McPherson, Smith-Lovin, & Cook 2001). In social network analysis, homophily can be measured by calculating an E-I Index (Borgatti et al., 2013).

Interdisciplinary Research:

> A mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice (Committee on IDR, 2004).

Network Constraint: A measurement of the degree of connectivity among a group of nodes with ties to Node X that can be used to reveal structural holes within a network (Borgatti et al., 2013). Network constraint is an inverse measure; lower constraint scores indicate the presence of more structural holes (Hanneman & Riddle, 2005).

Nodes: Points on a network graph representing an actor. (Borgatti et al., 2013).

Path: A set of one or more Edges connecting two Nodes in a network graph (Scott, 2013).

PI/Co-PI: An NSF abbreviation for Principal Investigator/Co-Principal Investigator. The Principal Investigator is the lead researcher who applied for and won an NSF award; there may be more than one Principal Investigator on a team, these members are called Co-Principal Investigators (NSF, n.d.).

Social Capital: "Resources embedded in a social structure that are accessed and/or mobilized in purposive actions" (Lin, 2001b, p. 29). In other words, social capital is the benefits and resources available to an actor through social connections.

Tie: The line connecting two nodes on a sociogram or network graph. An edge is indicative of a relationship between two actors, who are represented by nodes (Borgatti et al., 2013).

Undirected Network Graph: A graph in which the relationship between two connected nodes is equal, because they share a reciprocal relationship or are involved with the same activity. (Borgatti et al., 2013; Scott, 2013).

Vertices: See Nodes

Summary

With an increasing need for interdisciplinary research to solve complex problems in the world today, interdisciplinary research is often supported by funding agencies or organizations. With financial support, particularly from publicly funded institutions, stakeholders have a responsibility to measure outcomes of their investments. Social network analysis tools may provide stakeholders with useful information about the cohesiveness and multi-disciplinarity of research teams. Publicly available and archival data is used in this exploratory study to determine what social structural capital could be determined about a newly formed group of researchers by looking at co-authorship connections. This research also explores ways in which social network analysis tools can be used to characterize the interdisciplinarity or cross-discplinarity of those co-author teams.

This dissertation follows a five-chapter format. Chapter two contains a review of the literature and introduces readers to the history of CL and the development of a new area of scientific research. Chapter two also discusses Social Capital as the theoretical framework of this research and social network analysis as a mathematical way to make the Structural Social Capital visible to researchers and stakeholders. The historical use of Bibliometric and Scientometric data as proxies for first hand social network analyses are also discussed. Chapter three describes the design of the research. Specifically, chapter three details the three steps of the research process designed to the proposed research questions. Chapter four presents an analysis of the findings, arranged by the research questions. Chapter five discusses the implications of this research and provides recommendations for future work.

Chapter 2: Literature Review

This chapter provides a synthesis of literature relevant to this exploratory study of the use of archival, publicly available data to make visible the structural social capital and research domains of a small, newly formed, interdisciplinary group of researchers. The first section considers the emergence of scientific networks and the construction of scientific knowledge. A discussion of Social Capital, which is the theoretical framework of this study, comprises the second section of this chapter. The third section discusses the use of social network analysis processes and how these can be used to reveal the structural social capital inherent in a network. The fourth section of this chapter discusses applying social network analysis tools to co-authorship data.

Emergence of Scientific Networks

Scientific research, among many other things, is a social endeavor; even the solitary researcher shares her information with others in her domain and relies upon the work of others to inform and advance her own work. New knowledge creation relies on a combination of solitary or group research, and engagement with peers (McFadyen & Cannella, 2004). This communication among scholars has been referred to as an *invisible college* (Crane, 1972; Wagner, 2008). The term invisible college was first used to describe communities of scholars 17th century, when Robert Boyle and other scientists in England laid the foundation for the Royal Society of London by sharing their research (Wagner, 2008). Before the formal structure of a research institution, scientists in Boyle's invisible college shared their findings with one another, debated, and advanced their collective understanding (Wagner, 2008). Today, invisible colleges are informal networks of researchers whose expertise may span across disciplines or geography (Wagner, 2008). These informal scientific networks become a source of distributed knowledge that supports scientific advancement through the flow of information among the most productive members (Crane, 1972; De Solla Price & Beaver, 1966). Understanding the structure of informal scientific communities can reveal information about the flow of knowledge in the community. Graphs or sociograms can represent relationships between members of a research community, thus making the invisible visible. These types of representations can inform a research community about itself and provide data to stakeholders who may see brokering opportunities or areas in need of support.

Newly funded areas of interdisciplinary research share some commonalities with Boyle's invisible college. Newly funded interdisciplinary research areas take place outside of the formal structure of university departments. Also, researchers in a new interdisciplinary field lack a shared language of research and research methods, and lack a shared formal publication space. The work of CL researchers is an example of a newly created interdisciplinary community of researchers and thus presents a unique opportunity to observe the formation of a new scientific network.

Interdisciplinary Research

Interdisciplinary studies developed because some of the problems facing science were more complex than the academic departments into which knowledge has typically been divided (Ledford, 2015). The CL program studied here arose out of the recognition that understanding how people learn is itself interdisciplinary and that new technologies, how they can aid and support learning is a complex problem and cannot be tackled by a single discipline. The

National Academy of Sciences, National Academy of Engineering, & Institute of Medicine (2004) created this definition of interdisciplinary research:

> Interdisciplinary research is a mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice. (p. 2)

If one of the tasks of an organization is to foster interdisciplinary research, then those stakeholders have an interest in learning about the types and amounts of interdisciplinary research among their constituent scientists.

NSF recognizes that complex problems often require the knowledge and thinking of researchers from different fields. To meet these challenges, NSF funds centers to help build research communities, provide support to researchers and researcher teams, and bring together a wide variety of researchers and stakeholders. As such, “centers are a principal means by which NSF fosters interdisciplinary research” (NSF, 2011, p. 389). A CL program solicitation from 2012 included funding for a CL Resource Center. The role of the CL Center is to support grantees and disseminate their work; forge connections between researchers, investors from the private sector, organizations, and teachers; and create a national presence for CL by relying on current Cyberinfrastructure. Both CL program solicitations stressed that interdisciplinary teams would be needed to address the inherent complexity in the types of problems targeted by C:TE (NSF, 2010; NSF, 2011)[1].

---

[1] For more information on NSF’s CL program, see Appendix G.

Jacobs and Frickel (2009) argue that special evaluation criteria may not be required for "dramatically expanding interdisciplinary fields" (p. 52), because when an area of research develops rapidly and generates easily measured markers of success such as new conferences, new journals, and wins coveted award monies, evaluation criteria are easy to find. In most cases, the complexity of interdisciplinary research coupled with the fact that established evaluative models have not yet been developed creates challenges for stakeholders interested in assessment (Jacobs & Frickel, 2009). Social network analysis methods may provide stakeholders with tools to assess the growth of a research network and to understand some of the social structural capital that exists among researchers.

As Wagner et al. (2011) discuss, the process of determining individual research expertise is difficult, because researchers may not be working in the discipline in which they obtained their degrees and an analysis of researcher curriculum vita requires deep subject knowledge to correctly label expertise. The difficulty of categorizing researchers into a single domain presents a challenge for this research as well. Learning scientists, who make up 51% of the population studied in this research, often identify several domain areas in which they work and multiple areas of research interest (Yoon & Hmelo-Silver, 2017). Identifying research domain is relevant to this project, because a researcher who has collaborated with others outside his/her discipline suddenly has a connection to an entirely different research community. The idea is that who one knows matters and can influence the resources to which one has access is part of social capital theory.

Social Capital Theory

This section begins by providing some history on social capital and defining social capital; discussing social capital from individual and group perspectives; and explaining the structural components of social capital, with an emphasis on structural holes and brokering.

History. Social capital traces its roots back to Karl Marx's Theory of Capital, which defines capital as either a surplus that is received by a ruling class (i.e., a surplus created by selling the end-product back to the people who produced it at a higher cost) or as something that can be invested to create greater surplus (Lin, 1999; 2001b). Marx's theory, equal parts political economics and sociological commentary, influenced both economic theory and sociology. Pierre Bourdieu (1986) posits that the historical view of economic theory, tied as it is to capital (i.e., the accumulation of human labor), must be expanded to other types of capital to understand the "structure and functioning of the social world" (p. 241), since "even priceless things have their price" (p.242). Bourdieu's (1986) major contribution to the theory of social capital includes a consideration of cultural capital; cultural capital is the knowledge of the dominant culture in a society, which one must assimilate or master to succeed. Human capital is another type of capital that results from investments made by an individual (such as training or education) to increase one's status, position, or income (Lin, 1999). Social capital uses the basic economic theory of goods and services exchange and applies them to the benefits received in human relationships. Social capital also includes aspects of sociology; insofar as social capital looks at potential change that can happen on an individual or group level because of human relationships. Social capital, in contrast to human capital or economic capital, exists within a person's network, between individuals, rather than with an individual (Bourdieu, 1986;

Lin 1999, 2001b). There are two pieces to social capital, the human capital associated with the individuals in the network, and the structural connections between them.

Social capital is the potential or real value (i.e., capital) that can be realized because of human relationships (Bourdieu, 1986; Burt, 2005; Lin 1999, 2001b). Social capital is important, because the structure of social relationships impacts the flow of information, the way knowledge is spread through a group (Burt, 1992; Lin 1999, 2001b), and the influence held by individuals within a group (Bourdieu, 1986; Lin, 1999). Making the structural social capital of a group visible brings to light its potential in order that it can be acted upon e.g., potentially helping broker new strategic connections in a group to increase the flow of information through a group. The audience that has traditionally been interested in or privy to learning about the structural social capital of a group are those in administrative or stakeholder roles. It is this group that typically commands the resources to request a social network analysis and this group that is positioned to respond to the information.

Views on network density. Major works in early in social capital theory research seemed to be divided into two opposing camps: those who believed in the benefits of a densely-connected structure and those who argued that the value of a network is found in clusters or groups that were loosely connected. With researchers taking such strong stances, it would be easy to imagine having to choose one over the other. These views are not dichotomous or mutually exclusive as they seem and, by itself, network density is neither a positive nor a negative feature of a network. A dense network may generate trust (Bourdieu, 1986; Coleman, 1988; Putnam,2001); it can also impose such norms on a group that creativity is stifled or a group begins to think too narrowly (McFadyen & Cannella, 2004). Less dense networks may create ideal conditions for creativity and entrepreneurship (Burt, 1992; 2000; 2004; Granovetter, 1973), but there is less safety in such a situation.

Bourdieu (1986), Coleman (1988), and Putnam (2001) observed many examples of the benefits to individuals and communities with dense network ties. For example, Coleman (1988) cites studies done with diamond traders in New York, merchants in a Cairo market, South Korean student revolutionaries and the network of a single mother all as examples where a densely-connected network created an atmosphere of trust and security that enabled social support and the exchange of goods that could not happen in a more loosely connected network. The benefits of closed networks include increased trust, reinforcement of group identity and group norms and homogeneity. Stakeholders are often in positions to influence groups by bridging structural holes or increasing the ties to a group. It must be noted that these kinds of choices are not made because one network format is preferable to another on its own. Link most things, in social capital theory, context matters.

Weak ties and structural holes. Mark Granovetter's 1973 seminal article *The Strength of Weak Ties* didn't fully operationalize the term *weak ties*, but instead stated an assumption that a weak tie could be represented by a lack of intimacy, shortness of time known, or lack of emotional intensity between individuals. Granovetter contributed to Ronald Burt's work by introducing the concept of people acting as bridges across structural holes. Ronald Burt (1992) coined the term structural holes to talk about a gap between two individuals (or clusters of individuals) who have assets or information that might be useful to one another. Much of Burt's research has been about the benefits available to that person who can bridge the gap between two previously unconnected individuals. When the structure of social relationships can be shown to contain gaps (structural holes) e.g., where groups of people are not connected or are only thinly connected, those who manage to bridge these structural holes are uniquely positioned to benefit from information or enjoy influence within a network (Burt, 1992; 2000; 2004). Critiquing Everett Rogers' 1962 landmark book *The Diffusion of Innovations,* Granovetter (1973) pointed out that most of the diffusion research restricts itself to close ties; Rogers would later adopt Granovetter's suggestion to include weak ties in his research of rural Korea (Granovetter, 1973).

Those, who connect structural holes, are potential linchpins in the movement of information and ideas across a community (Burt, 1992;2000; 2004). Interdisciplinary research involves two or more areas of research; members of interdisciplinary teams play powerful roles and create bridges across structural holes that may exist between domains. For Burt, the advantages in network lie in the potential found in structural holes and with the individual who manages to create bridges across the gaps (Burt, 2000; 2004). Those who can bridge these structural holes or span the boundary between two groups are often those who can come up with more good, creative ideas, because they have been exposed to the language and processes of that other culture (Burt, 2004). It should be noted that merely bridging a gap does not increase social capital, there must be some value gained in the connection (Lin, 1999). Those researchers who span structural holes provide opportunities not only for collaboration between these groups, but for the exchange of knowledge among them. For example, Citizen Science is one class of CL project. Citizen Science, already an interdisciplinary area of research, is using web-based technologies to broaden citizen participation; this requires Citizen Science researchers to work with Computer Scientists who can develop the necessary infrastructure

(Crain, Cooper, & Dickinson, 2014). In successful collaborations, both the Computer Scientists and the Citizen Scientists establish working relationships and learn things about different domain. Newfound knowledge gained through collaboration aids Citizen Scientists in their thinking about future projects and connects them to other researchers in the Computer Science domain.

Burt eventually integrated what had been two dichotomous perspectives on social capital: the advantages of structural holes versus the advantages of densely connected networks. Both densely connected networks and networks with structural holes have certain advantages (Burt, 2000, 2005). Closed networks can generate trust among members and allow a group a deeper exploration of a topic or tradition. Structural holes are filled with potential to bring in new ideas, and as such can serve as hubs of innovation (Burt, 2004; 2005). Dense (closed) networks and sparse networks with structural holes complement one another in the larger social network, rather than being in opposition (Burt, 2005). Stakeholders with an awareness of the structural social capital of a group or organization may influence the structure to better serve the needs and purpose of the group (Cross, Borgatti, & Parker, 2002). Human relationships are inherently dynamic and susceptible to change; those with vision to see the potential in structural holes or areas that need more closure can better influence the direction of a group or organization (Cross et al., 2002). One way for stakeholders to change the structure of a group is through brokering activities.

Brokering. Brokering occurs when one agent forges a connection between two others, where a connection did not exist previously. The concept of brokering, like the theory of social capital, has been interpreted differently by various researches. The traditional view of brokering in social capital theory is that the individual forging the connection remains in the middle and benefits from that relationship by keeping the two entities apart (Burt, 1992, 2000; Gould & Fernandez, 1989). This orientation follows Simmel's (1950) triad concept of tertius gaudens, which is Latin for "the third who enjoys" (Obstfeld, 2005, p. 102). Tertius gaudens is a competitive model, in which the broker profits by keeping the two other agents apart; should the two entities meet the broker becomes superfluous. The work of a broker is not always one of exercising power. Krackhardt's (1999) research concluded that in some situations, those in brokering positions experience less autonomy to act, because they feel constrained by the normative pressures of the respective groups with which they have ties. Thus, neither of these perspectives on brokering are as conducive to the types of community building desired in interdisciplinary research.

Other researchers view brokering through a less competitive lens. Obstfeld (2005) proposed an alternate orientation to brokering, in which the broker consciously brings two parties together: tertius iungens. Tertius iungens also stems from Simmel's (1950) triadic theories, in which he referred to acts of brokerage that united individuals as "non-partisan" (p.146). Obstfeld (2005) translates tertius iungens as "third who joins" (p. 100) and refers to tertius iungens as an orientation that allows a person to help build more closure into a 'network. Obstfield's 2005 work shows that the brokering role can be non-competitive and create an entrepreneurial atmosphere within a group.

Brokering that results in innovation and brokering that results in competition are not mutually exclusive concepts and are in fact both desirable in many circumstances. Lingo and O'Mahoney (2010) synthesized these two orientations that brokering literature previously placed in opposition and both orientations are in fact enacted by those influential actors who are positioned at structural holes. Lingo and O'Mahoney's ethnographic study of Nashville music producers revealed that to move forward the complex problem of creating an album, producers found it necessary to keep certain individuals apart (the artists and record label managers) and bring people together (artists and other musicians). This integration of tertius gaudens and tertius iungens follows Portes' (1998) exhortation to look at complex social facts when looking at social capital.

Often, discussions of social capital focus on the benefits to individuals and groups that are realized via brokering or densely connected networks. It is important to realize that social capital does not necessarily equate positive outcomes. Portes (1998) cautions against a one-sided view of social capital that creates an "unmitigated celebration of community" (p. 22). Much research in social capital refers to the benefits experienced by a larger group of people through having strong ties (Coleman, 1988, Putnam, 2001), but it is also true that a densely-connected group may also discriminate against outsiders, serve to undermine the success of a group, and constrains individual freedom by requiring conformity (Portes, 1998). Social capital provides a lens through which the structure and inherent strengths of a community (at a given moment in time) can be analyzed. Understanding the structural social capital of a community and the human capital held by individual members affords stakeholders the opportunity to influence the structure to better serve the needs of the community at large, and therefore the stakeholders themselves.

By including the domains of the CL researchers in the analysis it was possible to see those researchers who have worked with researchers in other disciplines. Fields (2015) calls those researchers who co-author with researchers outside their own areas of expertise *cross-disciplinary brokers*. Stakeholders who can see structural holes in a network and identify those cross-disciplinary brokers are in positions to broker relationships across different disciplines to grow an area of research.

It's easy to conceptualize the ties among a small community in one's mind and to picture what connections might be beneficial to a group that do not yet exist. Tools are needed to help one comprehend larger networks. Social network analysis software and visualization tools provide the assistance needed to create sophisticated network graphs. These network graphs reveal the social structure of a group to which social capital theory can be applied.

Social Network Analysis

Stanley Milgram's well-known 1967 Psychology Today article *The Small World Problem* captured researchers' imaginations by showing that a surprisingly small number of people is required to connect any random two people. Beyond the entertainment factor of this story, Milgram's (1967) point was to show the existence of a mathematical structure to social relationships that has a broader impact. While Milgram's study contained some flaws (Van Raan, 2005), the small world phenomenon happens across all real networks and this measurement has become a mainstay of social network analysis techniques (Barabâsi et al., 2002; Newman, 2001, 2004).

Social network analysis is a way to measure and make visible the structural social capital in a group. Social network analysis techniques (or complex network methods) mathematically measure the social ties in a group or network. Different types of relationships yield different types of networks, friendship networks, kinship networks, and business networks are but a few examples. Most individuals belong to multiple networks. Sometimes there is overlap between networks, for example when two actors are connected in a business network and outside of the work environment have a friendship connection. When a network is displayed in a graph format, actors are represented by nodes. Members of a network being studied are referred to as actors and are represented as points on a graph by nodes (or vertices) that are connected by ties (or edges) that signify a type of relationship.

In social network analysis, nodes can be linked to different characteristics or attributes of the individual. Attributes are variables that show characteristics of the individuals and are called non-network data (Borgatti et al., 2013). Attributes can be either categorical or quantitative (Börner & Polley, 2014; Borgatti et al., 2013). Examples of categorical attributes included in this analysis are: Researcher NSF award numbers, researcher employment affiliation, and researcher domain.

Borgatti et al (2013) describe the ties between actors as relational states that can be described as either as “continually persistent” (p. 4) or events that mark discrete events. Relational states can be categorized three ways: Similarities, Relational Roles, or Relational Cognition (Borgatti et al., 2013). CL researchers with relational ties might be members of a PI/Co-PI team or may have been inspired to write together for other reasons.

Borgatti et al. (2013) make a distinction between the methods and uses of applied social network analysis and basic social network analysis. In basic social network analysis research uses dependent and independent variables and looks for correlations between them. Basic social network analysis research attempts to establish a network theory of X, where X equals one of the network variables. Applied social network analysis research is univariate and descriptive in nature and makes the structure of a network visible in order to inform decisions or actions (Borgatti et al., 2013). Applied social network analysis techniques were used here, because as an exploratory study, correlations are not part of the research questions.

There are two types of social network analysis research designs: Personal-network and whole-network (Borgatti et al., 2013). Whole-network, as the name implies tries to capture data from all actors in a network. Personal-network studies gather network information about individuals to gain insights to or an understanding about a group of people; personal-network studies do not seek connections between the individual study participants. Borgatti et al. (2013) mention that generally, gathering data for an entire network exacts costs on researchers and community members and results in less rich data than can be obtained by studying personal networks.

One of the first steps in social network analysis is to determine the boundary of the network to be studied. Sometimes networks are clearly bounded and sometimes the boundaries are less concrete (Borgatti et al., 2013). When considering how a network is bounded, researchers may adopt a realist perspective that looks only at formally recognized groups or a nominalist perspective that may study groups without formals borders; it is also possible to apply both realist and nominal perspectives to bound a network (Borgatti et al., 2013).

Social network analysis ideally involves a survey that includes all members of a group or organization. To get a complete picture of the group, a very high response rate is necessary. This type of analysis comes at a high participation cost for the community because a high return rate is necessary to create an accurate picture of a network. However, social network analysis techniques can be applied to secondary or existing data that represent communication or collaboration (Borgatti et al., 2013; Long, Cunningham, & Braithwaite, 2013). Some examples of types of existing, secondary data that lend themselves to social network analysis techniques: e-mail exchanges, listserv discussions, and lists of conference presenters or posters. Any of these can provide evidence of past collaborations to which social network analysis techniques can be applied. This study used citations found in bibliographic databases to discover co-authorship teams of CL researchers. Non-relational attribute data was gathered from publicly available online sources such as CV's and departmental websites.

The burden on a community during the data collection of social network analysis research can be greatly minimized if primary or secondary sources are used to provide data for whole-network designs. Bibliographic data and the increased number of information and artifacts found online are increasingly popular existing sources for whole-network research. This research used a whole-network design with data obtained from publicly available, existing resources. This data collecting method gathered information about all CL researchers, with minimal cost to community members and the researcher. The use of publicly available secondary data sources allows this research to be foundational for future longitudinal studies of CL researcher co-author collaborations. Drawbacks of this method include: The omission of published collaborations that are not listed in the databases searched; the exclusion of research collaborations that may be important or influential to researchers, but are not documented through published co-authorships.

Whether data is collected through interviews and surveys or secondary data, the connections between individuals can be revealed by applying social network analysis tools.

These tools can reveal features of the structural social capital of the group through the help of graphs, similar to genograms, called sociograms. In such a graph, each node represents an individual and each line that connects the nodes represents some form of a relationship. According to Burt (2005) a sociogram shows the “network residue” (p. 12) of human interactions. The lines, or ties, that connect the individuals’ nodes are indicative of a relationship; these ties can be one-sided or bidirectional (undirected) and they can be strong or weak. Once a network of relationships has been put in a format recognized by social network analysis software, mathematical principles can be applied (Lin, 2001a). Formulae and algorithms found in social network analysis software can be used to make visible constructs of Social Capital such as dense networks and structural holes.

Social network measures to be used. Social network analysis began with graph theory, Gestalt psychology’s emphasis on perception through understanding patterns, and early sociology and anthropology work in the United States and England (Scott, 2013). Social network analysis is a mathematical set of tools that allows researchers to look at relationships among a population (Borgatti et al., 2013; Scott, 2013). Social network measures can indicate social capital and the structural social capital of this research community was assessed by looking at both individual and total network properties, where appropriate.

Degree centrality, sometimes also called point centrality (Börner et al., 2005) is a measure of individual nodes that counts the number of edges or ties connected to a node (Bordons, Aparicio, González-Albo & Díaz-Faes, 2015; Börner et al., 2005). By using this local centrality, it is possible to rank the most connected individuals in a network from highest to lowest. Scott (2013) cautions that this number is always relative to a network. Borgatti et al.,

(2013) note that degree centrality is not a true centrality measure, but is included in that family out of tradition. A node with a high degree centrality is therefore well connected within a network and well-positioned to learn of new information as it travels through a network. A similar centrality measure used in this research, Eigenvector Centrality, calculates an Eigenvector for each node by weighting the degree centrality of its adjacent nodes (Borgatti et al., 2013).

In a highly-fragmented network, rather than having a large group of connected researchers, the network would consist largely of isolated clusters. Betweenness centrality is a measure that indicates the extent to which a member of a network lies on a path that connects different nodes within a network (Borgatti et al., 2013; Scott, 2013).

Cohesion is a measure of density in a network. Density factors can be calculated for different members of a group (Borgatti et al., 2013). By considering the attribute of researcher domain, it would be possible to see if a group of CL researchers is more connected than another (e.g., learning scientists versus computer scientists).

Social network analysis and-co-authorship networks. Newman (2001) argues that the co-author relationship is significant enough to be counted as a connection, since researchers become well-acquainted with one another through the collaborative work of writing a paper together. Co-authorship relationships among scientists can be used to model social networks (Yin et al., 2006); this structure lends relations can be mathematically measured is collaboration strength, which can be represented by the number of times one author has partnered with another to write a paper (Börner et al., 2005; Newman, 2004).

Prior application of social network analysis techniques to co-authorship networks. There are three ways to establish the population to be studied in a co-authorship analysis: Studying a known population, creating the population from an exhaustive database search, or beginning with a set of articles. De Solla Price and Beaver's groundbreaking 1969 research was the first to study co-author relationships among a bounded group of researchers Since then, it appears studies using a known population of researchers are slightly less common (Börner et al., 2005; Milesi, Brown, Hawkley, Dropkin, & Schneider, 2014; Rodriguez & Pepe, 2008; Yin et al., 2006) than those created from databases (Barabâsi et al., 2002; Bordons et al., 2015; Glänzel & Schubert, 2005; Li, Liao & Yen, 2013; Liu, Bollen, Nelson & Van de Sompel, 2005; Newman, 2004). Xian and Madhavan (2014) began their analysis with over 24,000 engineering papers.

Bibliometrics and Scientometrics

History of bibliometrics and scientometrics. Bibliometric data has been commonly used to help paint a picture of and understand the growth and development of science for over 60 years. In 1960, Eugene Garfield founded the Institute for Scientific Information and began the Science Citation Index in 1961. Garfield (1979) had the foresight to understand that the bibliometric data from scientific literature could be important for understanding the landscape of science as well as understanding where it is headed. The Science Citation Index allowed researchers to quantify scientific research on a scale that was not previously possible (Van Raan, 2005) and allowed researchers to begin examining co-authorship networks. Derek De Solla Price used the massive amounts of data available in the Science Citation Index to begin to formulate laws governing scientific publications and ways to understand the topology of science (Van Raan, 2005). In 1965, De Solla Price published research demonstrating the exponential growth of science. Works published by De Solla Price from 1960-1970 forged the way for Scientometric research.

Scientometrics and Bibliometrics are closely related quantitative methods that use publications as the unit of analysis. These two statistical approaches share commonalities in their interdisciplinarity and their shared use of similar sources of data (Mooghali, Alijani, Karami, & Khasseh, 2012). Bibliometrics and scientometrics both look at the growth of science, concern themselves with the establishment of disciplinary boundaries, and measure output indicators (Godin, 2006; Leydesdorff & Milojevič, 2015; Van Raan, 2005). Though both Scientometrics and bibliometrics analyze a discipline's scientific literature with the goal of understanding something about the dynamics and growth of a field, most researchers make distinctions between the two terms. Bibliometrics, the older of the two terms, is concerned mostly with output measures of literature: the influence of an author, article, or journals by counting the number of publications, the number of citations a publication has received or co-citation analysis (Ding, 2011). Journal impact factors stem from Bibliometric research. Hess (1997) defines scientometrics as the "quantitative study of science, communication in science, and science policy" (p. 75). In addition to looking at the development of research fields as seen through artifacts of scientific activities (Hood & Wilson, 2001), measuring scientific impact for policy makers and funding agencies (Leydesdorff & Milojevič, 2015; Van Raan, 2005), using visualization tools to create domain maps based on scientific publications (Boyack, Klavens, & Börner, 2005), Scientometrics is also part of the sociology of science (Hood & Wilson, 2001). Scientometric research also includes the use of sophisticated datamining techniques used to gather attribute data for co-authorship networks by categorize research into different taxonomies and researchers into domains (Xian & Madhavan, 2014). Co-authorships represent a collaborative relationship. Co-authorships across a community of researchers or a discipline form a network of relationships that can be statistically analyzed to reveal structural elements, such as which researcher might act as brokers within the community of researchers being studied (Newman, 2004).

In the 1980's, the sociology of science began to focus on qualitative research and scientometrics concentrated on impact indicators relevant to policy makers (Leydesdorff & Milojevič, 2015). Thirty years later, the quantitative data of Scientometrics can be used to point to sociological aspects of science, specifically the co-construction of knowledge and impacts of co-authorship. (See: Bordons et al., 2015; Börner et al., 2005; Boyack et al., 2005; Ding, 2011; Glänzel & Schubert, 2005; Li, Liao & Yen, 2013; Liu et al., 2005; Newman, 2004; Rodriguez & Pepe 2008; Yin et al., 2006). Historians, sociologists, research evaluators, and policymakers all have an interest in the development of research fields. Over a longer period the combination of bibliographic and other publicly available data can mark the development of a field by identifying the major researchers and their areas of expertise, the researchers with whom they publish, and the extent of interdisciplinarity in the co-author relationships. Such an analysis reveals information about not only about the social relationships between researchers, but can also depict the structure of scientific knowledge (Newman, 2004). Boyack et al., (2005) used co-author relationships to the connections among many research domains

One of the drivers behind bibliometrics and scientometrics is to measure the impact of a piece of research on science. Van Raan (2005) discusses the fact that there is no single methodology that has been agreed upon for measuring science impact; in fact, he argues that a single method is not necessarily desirable. What matters, according to Van Raan (2005) are indicators: “An indicator is a measure that explicitly addresses some assumption” (p. 22). In this research, co-authorship is an indicator of collaboration. Van Raan (2005) also posits that indicators standing alone are irrelevant, they must be situated in the research and they must be stable over time and “not a noisy set of measures” (p. 31).

The domain in which researchers are publishing must be taken into consideration when looking at centrality measures, since different areas of research have different cultures and norms of publishing. The structure of the co-authorship network seems to be dependent on the type of scientific research (applied versus theoretical) in each field (Barabâsi et al., 2002; Bordons et al., 2015; Newman, 2001). Bigger scientific projects require large teams and shared resources and tend to produce papers with a larger number of authors than a theoretical field like mathematics, which does not require a large team. Nevertheless, co-author collaborations appear to be on the rise, making co-authorship a promising area to begin looking for the evidence of research collaborations (Rodriguez & Pepe, 2008).

Co-authorship and social capital. Statistical tools can be applied to co-author relationships within a discipline or among a select group of researchers to reveal the social structural capital within that group. A co-authorship analysis is one way to begin looking at the social capital that arises in the process of co-writing and article. Co-author relationships throughout a discipline or a research group form co-authorship networks (Newman, 2004). Co-author relationships represent social capital within a group of researcher. Newman (2004) states "The co-authorship network is as much a network depicting academic society as it is a network depicting the structure of our knowledge" (p. 5200).

Trust is required for teams to work and publish together. In the world of research, scientific reputation is valued and cultivated through one's work; the choice of co-author potentially impacts one's career. Interdisciplinary co-authorships serve a bridge across disciplines, while publication with a senior researcher can lend some credibility to a newcomer to a field. Collaborative work necessarily builds a certain amount of relational capital, both through the research and the process of co-writing (Laudel, 2002; Li, Liao & Yen, 2013).

The construct of trust surfaces repeatedly in the social capital literature about closure (Coleman, 1988; Putnam, 2001) and about brokering (Burt, 2004, 2005). The collaboration required to research and co-author a paper is indicative of a certain level of trust (Laudel, 2002; Li, Liao & Yen, 2013). For example, looking at the number of times two researchers have collaborated provides an indication of the level of trust between them. While trust was not directly addressed in this research, the PI/Co-PI relationship is indicative of trust, as is repeated co-authoring. Writing a paper with another person is the relational event this study analyzed. A second piece of relational cognition that can be difficult measurement is the flow of knowledge or information throughout a network; flow of information is generally an outcome of a relationship or an exchange (Borgatti et al., 2013; Burt, 2004).

Co-author networks in established fields tend to display the same small world properties as real communities (Barabâsi et al., 2002; Newman, 2001, 2004). Co-author networks in established fields tend to be scale free and display a power law, which means a small number of individuals are highly connected with many members not highly connected (Rodriguez & Pepe, 2008).

Co-author relationship limits. Co-authorship can be used as an indicator or proxy measurement of collaboration, though it is imperfect (Katz & Martin, 1997; Laudel 2002). No one would argue that two people who publish together are not collaborators, but that relationship doesn’t represent the entirety of collaboration behind the research. Major research projects require the efforts of scientists, research assistants, and lab technicians, and those projects sometimes involves shared equipment. CL projects that designed to bring ideas or tools to scale require work with schools, teachers, industry, and organizations; the authors are simply the most visible subset of a research team in this measure (Katz & Martin, 1997; Laudel, 2002). Furthermore, there are many other ways to collaborate with a colleague that leave no trace, even a conversation between two researchers who are not formally collaborating can have an impact on a research project (Katz & Martin; Laudel, 2002). "The bibliometric indicator 'co-authorship' is systematically biased against some collaborative practices” (Laudel, 2002, p.4). The relationship required to share equipment with another group of researchers may bridge a valuable structural hole in a researcher’s network, but that relationship may not be shown through the publication that results from that collaboration. Finally, honorary authorships create a situation where not everyone who is listed as a co-author has necessarily participated in the writing of a paper (Laudel, 2002; Glänzel & Schubert, 2005). Knowing these limitations, the analysis of bibliographic data still can provide insights into the social structural capital of a research community and help researchers better understand scientific collaboration (Ding, 2011). Despite these limitations, co-authorship data is a logical place to begin to chart the growth of collaboration among a group of researchers or in a discipline, because this data is well documented it is easy to construct graphs that show how a co-author network grows and evolves over time (Newman, 2001.)

Summary

This chapter reviewed the literature in the three main content areas that support this research: Social capital; Social network analysis, bibliometrics/scientometrics. Researchers who study social capital among scientists and engineers coined the term scientific and technical capital to talk about the combination of human capital, social capital, and knowledge, including tacit knowledge that moves with a researcher from post to post (Dietz, Chompalov, Bozeman, Lane, & Park 2000). All knowledge is socially constructed and flows through networks of people (Borgatti & Cross, 2003). In order for this knowledge to flow through a network, a community must meet certain criteria: Members must know who knows what; the knowledge held by others in the community must be useful; it must be easy for members of the community to connect with one another (Borgatti & Cross, 2003). Layering researcher expertise into scientific network graphs allows for the identification of cross-disciplinary brokers and can make visible those areas in which it appears there are not yet connections. Because co-authorship is evidence of a collaboration and an indicator of knowledge flowing through a community, the use of bibliographic data can be a nonintrusive way to find connections within a known community of researchers. Stakeholders in research projects, who understand the structural social capital of research communities, may find this supports an evaluation of their projects. In the interests of supporting interdisciplinary initiatives and understanding the growth in specific areas of research, applying social network analysis tools to bibliographic and other publicly available data has potential to show the social capital strengths and weaknesses of a community of researchers.

Chapter 3 discusses the methods used and the theoretical orientation that informs this research to answer the study's research questions.

Chapter 3: Methodology

This study explored the feasibility of using publicly available data to identify the key domains in which a bounded community of researchers work and to make visible the social structural capital existing in the community via co-authorship ties. The choice to use publicly available data was an intentional aspect of this research design. Those researchers who apply for and win highly competitive, national grants are top-level, ambitious, and extremely busy; conducting a true social network analysis of their research networks would place a burden on the community and questions coming from an unknown graduate student would not likely yield sufficient results. Using publicly available data has the additional advantage that this research could be replicated. A further benefit of the hands-on, artisanal method of collecting data employed in this study may reveal nuances that inform larger scale efforts to collect similar kinds of data. In this study, social network analysis tools were applied to bibliographic and other publicly available data to show the early structural social capital inherent in a network of researchers. What follows is a discussion of research design, data collection, processing, and analysis; ethical considerations; and limitations of the study.

Restatement of Purpose and Research Questions

Funders and stakeholders in all areas of research have an interest in understanding the composition of the research communities they support, as well as gaining an understanding of the growth, impact, and diffusion of ideas and research within those research communities. The community of researchers funded by the National Science Foundation's (NSF's) Cyberlearning and Future Learning Technologies program and the past program Cyberlearning: Transforming Education studied here serve as an example of a type of community that lends itself to this sort of research. Of potential benefit to the CL community specifically, is the creation of a foundational understanding of members of that research community and ways in which members are connected via co-authorship networks. The matrices and graphs created in this research provide baseline measurements that illustrate relationships among the entire network of CL researchers. Results of this study may provide insight into the structural social capital of CL researchers and may also provide a point of reference for looking at growth, impact, and diffusing in 5-10 more years by answering the following questions:

- RQ1: What underlying social capital structures can be determined about a group of researchers from bibliometric data and other publicly available existing data?
- RQ2: What are ways social network tools characterize the interdisciplinarity or cross-disciplinarity of co-author teams?

Research Design

The theoretical framework underpinning this research design was social capital. Social capital posits that the ways in which individuals connect with one another influence the ways knowledge flows through a network (Burt, 1992, Crane, 1972; De Solla Price & Beaver, 1966; Lin 1999, 2001b). At the beginning of any new area of research, such as CL, information about the communication patterns of researchers could potentially aid stakeholders in understanding the directions in which the field is developing.

A post-positivist epistemological perspective underlies this project. Post-positivists believe that such a thing as reality or a truth exists, but that research can only make imperfect attempts to attain it (Phillips & Burbules, 2000). Therefore, whatever insights this research may provide about the methodology used or the social capital highlighted in the results, is only one view of the social capital in the CL community.

Co-authorships networks are a type of social network, to which social network analysis tool can be applied (Bordons et al., 2015; Börner et al., 2005; Boyack et al., 2005; Ding, 2011; Glänzel & Schubert, 2005; Li, Liao & Yen, 2013; Liu et al., 2005; Newman, 2004; Rodriguez & Pepe 2008; Yin et al., 2006). By applying social network analysis techniques to data collected from WoS, Google Scholar, and publicly available websites, it was possible to address the research questions in this study.

Setting and Participants

This research uses existing, publicly available data and as such, there were no active participants. Given the public nature of the data sources for this research and that no humans were interviewed for this study, this research was approved as exempt from Institutional Review Board (IRB) oversight. Data was collected about those NSF Principal Investigators (PIs) and Co-Principal Investigators (Co-PIs) from the Cyberlearning: Transforming Education (C:TE) program and the Cyberlearning and Future Learning Technologies (C:FLT). Most of the 230 CL awards were given between 2011-2015; five awards from 2007-2010 fell into the category of CL and were included in this study. The PI/Co-PI population was comprised of 411 researchers. The CL program distributed awards to grantees in 40 of the 50 United States (CIRCL, n.d.-b.).

Data Processing and Analysis

Data processing and analysis took place in three parts: Data cleaning and sorting; applying social network analysis tools and visualizing the network; attribute coding; adding attribute data to the network visualization. Tools used in these four steps included: Endnote, Microsoft Excel, Microsoft Access, a Python program, and UCINet (Borgatti et al., 2002)).

Data cleaning and sorting. Endnote was used as a collection and storage tool for the bibliographic records found in WoS, because it is a tool that works well with both WoS (Endnote is a Thomson Reuters product) and Google Scholar. Since format of author names differs between WoS and Google Scholar, the 182 Google Scholar records were edited to conform to the WoS format. The data from Endnote was then downloaded into a CSV format; Microsoft Access software was used to remove duplicate entries, resulting in 4,006 unique articles. Next, a CSV file of the author names was run through a Python program that extracted all of pairs of CL grantee names from each citation record and counted the number of times the pair collaborated on papers that were found among the 4,006 articles.

After finding the pairs of co-authors, the number of researchers was reduced to from 411 to 191. The extraction of pairs of CL co-authors from the list of all authors created a valued edge list that served as the basis for the calculations run by the social network analysis software UCINet (Borgatti et al., 2002). Social network analysis techniques were applied only to this subset of 191 researchers.

Applying social network analysis tools. When looking at relatively small networks, it is possible to draw a sociogram without the aid of software or other tools. If a network is larger, creating matrices provide the information needed for social network analysis tools to generate network graphs. Data often begins in what is called a case-by-case matrix, also referred to as a square because the data in the row and column headings is identical. These square case-by-case matrices, or adjacency matrices, are the core of creating a network graph; adjacency matrices show who (or what) is connected to whom (or what) and at what frequency (Scott, 2013). The valued edge list of 191 pairs was uploaded to UCINet (Borgatti et al., 2002) and converted into an adjacency matrix. This matrix was then uploaded into NetDraw to create a visualization of the network.

Sociograms or maps of network relationships can be visualized with network graphs (Börner & Polley, 2014). Network graphs are made up of nodes and ties (Borgatti et al., 2013). In this research, nodes represent the individuals in the network, while the ties are indicative of their co-author relationship. The valued edge list of co-author pairs established that a pair of researchers had a co-author relationship and should therefore be connected by an edge on the graph. The number of co-author collaborations determined the strength of the tie between the nodes and was represented by the width of the tie between nodes. The strength of the co-author relationship can be demonstrated by the thickness of the line between two nodes. Co-authorship graphs are undirected, this means there is a symmetry in the relationships among co-authors (Liu et al., 2005); unlike a relationship where person X might discuss ideas with person Y, but person Y may prefer to confide in person Z, a co-author relationship represents a symmetrical relationship.

Formulae. The selection of mathematical functions used to analyze network data is contextual and depends in part on the type of information flowing through a network. Centrality measures are used to reveal some of the basic network structures. Borgatti (2005) talks about using appropriate centrality measures for appropriate networks, because every centrality measure is based on an implicit assumption about the flow of a network. Flow through a network depends on what it is that is flowing through the network. In this case, we are looking at information, knowledge and ideas about what works and what doesn't by a research community. In the case of networks like this one, where information is flowing between the nodes, Borgatti recommends calculating Degree Centrality, Closeness Centrality, and Eigenvector Centrality. These measures, as well as constraint, cliques, homophily, and brokering were analyzed in this study.

Degree centrality is a simple measure that counts the number of ties connected to each node (Borgatti, 2013, Scott, 2013). This measure is always relative to the network. In some ways, degree centrality can be regarded as a popularity measure. By adding attribute data to the CL network data, it was possible to see the domains of a node's co-authors, thus shedding some light on cross-disciplinary publishing within the CL community. Eigenvector centrality is a more sophisticated type of degree centrality. Eigenvector number is a weighted centrality measure that is based on the number of ties one node has to others, but when calculating Eigenvector centrality, all nodes are not created equal. Eigenvector calculation gives each node in the network a weight, based on the number of nodes to which it is connected (Borgatti et al.,2013; Scott, 2013). This is an appropriate measure when thinking about contagion – a person is at greater risk of catching a disease or hearing about a great idea if she is connected to others who are well-connected (Borgatti, 2005; Borgatti et al., 2013). Related to degree centrality and Eigenvector numbers is closeness centrality. The closeness centrality calculation is a sum of one node's distance from all other nodes (Borgatti, 2005). A low centrality number indicates that a node is positioned to be exposed to information sooner than a node with a higher centrality number.

Constraint is a measure of how many of a node's ties are connected with one another (Borgatti et al., 2013, Scott, 2013). The constraint formula measures the degree to which those nodes adjacent to node X have connections with one another and thus is a way of calculating and making visible structural holes within a network (Borgatti et al., 2013; Hanneman & Riddle, 2005). Following Burt's (1992) theory of structural holes, if there are many connections among those nodes tied to Node X, then Node X can's options are limited. Krackhardt's 1999 research also revealed that a densely-connected network could constrain one's actions. Network constraint can be calculated both for directed and undirected networks (Hanneman & Riddle, 2005). Network constraint is an inverse relationship, the smaller the node size, the fewer ties there are among those nodes connected to Node X. Therefore, a low measure of constraint can be viewed as an indicator of the presence of structural holes; the lower the constraint number, the greater the opportunity for a researcher to broker connections across structural holes and to be exposed to new ideas. A node's constraint number can range from zero to one, though in smaller networks (such as the one in this study) it is possible for constraint numbers to be above 1 (Hanneman & Riddle, 2005). Therein this study data was analyzed using the UCINet's (Borgatti et al., 2002 Structural Holes tool and applying the Ego Network Model that uses a method suggested by Burt (1992) of looking at only links two steps away from each node when calculating constraint.

Cliques, as defined by social network analysis, are maximally connected groups that must contain at least three nodes (Borgatti, 2013). In simple terms, a clique is a group of nodes, in which each node in the group is connected to every other node in the group. It is possible for cliques to overlap, this is one reason that as clique size is increased, the number of cliques in a network are reduced. In this study groups of cliques were assigned the non-relational attribute of award number to obtain a sampling of the composition of award teams that might exist among co-author collaborations.

The homophily formula uses attribute data to partition a network into different groups and then calculates the number of like nodes to which and given node is tied. In the case of the CL network, the data was partitioned by domain. Homophily can be calculated at both the individual or node level or network level. In both cases an E-I score is given. E-I scores range from 1 to -1; a score of 1 indicates a perfectly heterophilous network, 0 indicates a perfectly balanced network, and -1 indicates a perfectly homophilous network (Borgatti et al., 2002). This measurement was selected to measure the cross-disciplinarity of co-author teams (as represented through domain variation among co-author teams).

Finally, the measurement of brokering was used called G&F, after the expert researchers in brokering, Gould and Fernandez. UCINet (Borgatti et al., 2002) bases its G&F brokering calculations on Gould and Fernandez's 1989 research and their model of five different brokering roles: Coordinator, Gatekeeper, Representative, Consultant[2], Liaison. These five roles describe the relationship of a single node bridging two other nodes and encompass all possible combinations. Attribute data for researcher domain was included in these brokering calculations to determine which role(s) apply to each node. The node in a coordinator role is positioned between two like nodes; this is a homophilous group. A consultant position finds a node sandwiched between two other nodes that belong to the same group, but a different group than the consultant. In the liaison role, the brokering node

[2] The term consultant is used in UCINet reports and is the term used in this dissertation. Gould and Fernandez used the term Itinerant in their 1989 paper.

connects two nodes that are different from the broker, and also different from each other; this as a completely heterophilous group. Both the gatekeeper and representative roles include the brokering node and one other node from the same group, with the third node having membership in another group; because co-author relationships seen in this study are undirected, the gatekeeper and representative roles are effectively identical. Calculating which nodes occupy which brokering roles allows stakeholders to see what types of brokering potential exist within a network. And in a co-author network this is another way for stakeholders to look at how (and if) researchers from one domain are collaborating with researchers from another.

Gould and Fernadez's (1989) five brokering types describe the potential of a node's position within a network, rather than actual work that is done by the node. Every node that is the only connection between two other nodes falls into one of these five categories. It is possible for a node to be assigned to multiple brokering roles. These brokering roles are counted by instance; it is, therefore, possible for node X to have three different liaison connections and 2 different gatekeeper connections.

Attribute coding. Identifying the domain in which a researcher works was one of the key pieces of information needed to answer RQ2, which asked about identifying the cross- or interdisciplinarity of research teams. To classify this group of researchers into their respective domains, information gathered during the initial data collection phase of this research was used: Job title or departmental affiliation, last academic degree awarded, and stated research interests, where available. Using a technique highlighted in research by Schlager, Farooq, Fusco, Schank, and Dwyer (2009), areas of interest on the network graph were selected and attribute data was coded for those groups. This resulted in domain attributes being assigned to 114 researchers; these researchers were those in the main component of 60, and two smaller components of 12 and 18 researchers, plus 20 researchers from 10 randomly selected dyads and one 4-member clique. These 114 researchers were each assigned a single research domain that reflected their work.

Klavans and Boyack (2009) created a taxonomy of 16 research domains in their research mapping scientific communication[3]. The Klavans and Boyack (2009) list contained to many elements to be applied to this sample, but provided an example and inspiration for the domain codes used in this research. Three pieces of information were evaluated to assign domain affiliation to each researcher: Departmental affiliation, Domain of last degree obtained, Research areas of interest. If the department and degree were a match, the researcher was assigned to that matched domain. Sometimes a researcher would not have a clear departmental affiliation that fell in line with the way the academy has organized knowledge (because the researcher is working outside of the academy, or perhaps the director of an interdisciplinary center within the academy), in these cases, the researcher was assigned the domain of the last degree s/he obtained. Those cases in which departmental affiliation and degree domain did not match, which was common among these researchers, the research interests were consulted, if available to look for the closest match between the two. If no research interest information could be found to reconcile the two the departmental information was selected, unless the researcher's degree was Learning Sciences. This first cycle of coding resulted in the assignment of 15 different domains. A second pass consolidated the

[3] See Appendix B.

15 domains into the following seven codes: Learning Sciences, Computer Science, Human Computer Interaction, Interdisciplinary Studies, Science/Math/Engineering, Cognitive Science/AI, and Humanities[4].

There are several hurdles in assigning a domain to CL researchers that are not easily cleared. First there is the issue of interdisciplinarity. While CL itself is interdisciplinary, it is also possible for researchers themselves to have incorporated knowledge from multiple disciplines into their work. Also, the dynamic work of top level researchers can mean that over a period of time, they identify more closely with different domains. While the coding method used in this research may not neatly clear the hurdle of identifying researchers in the way the researchers themselves would, it nevertheless allows a proof of concept of this design.

Adding attribute data. When creating network visualizations designers can add attribute data to the network data (as mentioned earlier), creating a richer understanding of the network being studied. Certain design elements encode quantitative data (such as node size or edge width), while others encode qualitative (such as node hue or shape) (Börner & Polley, 2014). Table 1 shows the attributes and the encoding elements assigned to them that were used in network graphs in this research. The number of times researchers published together was included as a numeric (quantitative) value represented by the thickness of the tie between them in the network graph. Qualitative categorical attributes about the domain in which researchers work and in what type of institution they are employed were represented through the color and shape of the nodes, respectively.

[4] See Appendix C for the coding rubric.

Table 1

Attributes and Their Coding Elements

Tie/Node	Attribute	Encoding Element	Type of Data
Node	Area of Expertise (Departmental Association, Project Tag, Population Focus of the Award	Hue (tint)	Categorical
Node	Employment type (University, Research, Private, Non-Profit	Shape (tint)	Categorical
Tie	Strength of tie	Thickness	Numeric

Network graphs created from this research:

- Simple network graph, with ties weighted for number of collaborations between nodes, using encoding elements above to add more information to the network.
- A graph of the large component containing attributes of each author.
- A graph containing the large component, the components with 12 and 18 nodes, as well as 10 randomly selected dyads.
- A graph of cliques found in the network, with attributes.

Data Collection

Data for this research was collected primarily from six publicly available sources: NSF Award Search, Deep Insights Anytime Anywhere (DIA2), Thomson Reuters Web of Science Core Collection (WoS), Google Scholar, organization or department websites, and personal websites. This combination of resources proved necessary, because it was not possible to gather all pieces of information needed for this study from a central location. Additionally, some resources provided data in more usable formats than others.

A deliberate choice early in the planning of this research limited the use of Google Scholar as source for only award number searches, restricting the number of results obtained from the Scholar search engine. Google Scholar provides a more comprehensive picture of an author's work than an indexing/abstracting database like WoS because it is not restricted to a set of journals; instead, the search engine scours the Web for every scholarly publication it can find. Harzing and Van der Wal (2008) argue for the use of Google Scholar in tenure analyses for those reasons mentioned above. As part of their critique, they list the known issues with Scholar that prohibited broader use of the resource in this dissertation. The biggest disadvantage to using scholar for this work was Scholar's inaccurate, incomplete, or non-standard citation formats (Harzing & Van der Wal, 2008). A second issue that became apparent in the data collection process is the inability to easily download citations from scholar.

Data collection took place in two parts: First, establishing the list of CL grantees and collecting data about their awards, places of employment, and degrees (to establish in which domain they work); this qualitative data established the attributes that would be assigned to individual researchers and used in the analyses. Second, collecting citations from publications authored by each awardee between the years of 2009 and 2015 provided the relational information that would be used to illustrate the network connections among these researchers.

Grantees list and grantee data collection. The NSF program reference code for Cyberlearning (8020) was entered as an advanced search option within NSF Award Search database. This search returned 228 awards, with 411 PIs/Co-PIs. Award Search offers several download options; the data about these awards was downloaded into a Microsoft Excel file. Information relevant to this project that was retrieved from the Award Search download: Award Number, PI/Co-PI Names, Award Title, Award Type, and Abstract.

After completing research in the Award Search database, the next step was to gather data about PIs/Co-PIs places of employment (to help establish domain and type of institution) and the title of their last degree (also to help establish an area of research). Because many researchers have common names, the DIA2 data mining tool was the initial resource for researcher employment affiliation. Next, a web search was conducted for all 411 researchers to collect information to help situate the researchers in a domain (e.g., Verifying place of employment and departmental affiliation if applicable; title and domain of final degree; link to CV, where possible; research interests, where available). The availability of this information varied by researcher. All data collected in this first stage of research was collected in a Microsoft Excel spreadsheet. Table 2 below lists the types of data that were included in this first round of data collection and the rationale for inclusion.

Publications and co-author data collection. The second part of data collection focused on obtaining the network data. WoS Core Collection and Google Scholar was used to obtain names of those who have co-authored papers, conference proceedings, or books with CL researchers (PIs/Co-PIs) between 2009 and 2015. This date range begins the year after the publication of *Fostering Learning in a Networked World* and ends in the final year of C:TE funding. Data found in WoS Core Collection and Google Scholar provided the co-authorship names and connections that form the ties in the network graph of CL Co-Authorship. This study looked only at those writing collaborations between members of the CL grantee community.

Table 2

Data Collection Part One

Researcher Data Points	Obtained from	Reasoning
Award Number	NSF Award Search	The award number indicates a project. Everyone on the project has the same award number. This can be used to help determine multi-disciplinarity within a team.
Award Title	NSF Award Search	For Reference Only
Type of award	NSF Award Search	(e.g., DIP, EXP) Award type[5] is tied to funding; bigger projects are more likely to have more collaborators.
PI/Co-PI Names	NSF Award Search	These are the actors of the research network
PI/Co-PI Institutional Affiliation	DIA2	For locating CVs or data on organizational websites
Domain of last degree awarded	Personal or Org Website	Masters & PhD. Helps situate a researcher within a domain
Departmental Association	Personal or Org Website	Helps situate a researcher within a domain.
Departmental/Personal Website	Personal or Org Website	For Reference Only
Research Interests	Personal or Org Website	Helps situate a researcher within a domain

Web of Science (WoS) Core Collection database combined with Google Scholar search engine provided the bibliographic information needed to examine the co-author relationships among CL grantees. Both WoS and Google Scholar were accessed to obtain a more complete

[5] See Appendix G

picture of co-author collaborations in the CL community. WoS Core Collection combines seven of Thomson Reuters' indexes to create a multidisciplinary citation database that provides bibliographic information extracted from high impact journals in the Humanities, Social Sciences, and Sciences, while also indexing conference proceedings and books (Thomson Reuters, 2014). By contrast, Google Scholar is a search engine that indexes individual scholarly papers across the web, as such, results from searches in Google scholar were anticipated to uncover citations that did not overlap with the more restrictive WoS search results.

In WoS Core Collection database, author searches were conducted for those articles written between 2009-2014. The author search format in WoS is: last name, first initial(s). This query can create very large search results with common last names such as Chen, Zhou, Martin, or Singh. In the planning stages of this research, it seemed that the WoS Distinct Author Identifier search feature would allow authors to quickly and easily be identified. The data collected in the first stage of this research proved invaluable to the process of discovering the work of specific authors. Curriculum Vitae often showed an author's middle name or the usual format in which she publishes, and was sometimes used to validate that the researcher was the author of a publication in question. In the WoS Core collection, it was possible to narrow search results by researcher affiliation and/or research domain; it was often necessary to consult the data collected in the first phase of this research to identify the correct author in WoS. To save time, all publication citations attributed to Cl researchers were saved and download, even if the citation reflected a single author publication that ultimately would not be included in this study of co-author collaborations. Afterward all researcher names had been searched in the WoS core collection, close to 4,500 articles citations had been downloaded.

Google Scholar was used to discover any additional awards that specifically reference a CL award that may not have been included in the WoS results. Google Scholar searches followed this format: “National Science Foundation” + NSF award number. This format ensured that any results would be highly relevant, yet may include references that were not included in the tightly curated list of resources indexed in WoS. Only those publications referencing a CL award number and written by more than one author were included in those articles saved from Google Scholar searches; 182 articles meeting these criteria were downloaded.

Data gathered from WoS and Google Scholar was saved to the bibliographic management software EndNote. Placing the publication records in EndNote allowed the WoS and Scholar records to be formatted identically and then downloaded into a CSV file for analysis. See Table 3 for the four data points that were included in the EndNote data and the rationale for including them.

Ethical Considerations

Generally, network analysis focuses on individuals and the relationships between them. When people are asked about their relationships, they reveal information about themselves and about others (Borgatti et al., 2013). Unlike a traditional network analysis, this study applied network analysis techniques to existing data obtained from publicly available sources. (The National Science Foundation’s Award Search database, Deep Insights Anytime Anywhere (DIA2) datamining tool, University Websites or publicly posted Curriculum Vitae, Google Scholar, and bibliographic information from the Web of Science Core Collection database.) All the people for whom data was collected are individuals who applied for and received funding from the

United States' National Science Foundation; researchers understand that transparency is a requirement of receiving federal funding. Furthermore, the culture among research scientists is to publish publicly to share information with other scientists; having one's work seen is desirable. As part of the IRB exempt status of this research, the data was anonymized for publication and no names appear in the network graphs. The results of this study are discussed in the next chapter.

Table 3

Data Collection Part Two

Data	Rationale
Authors and Co-Authors	Needed for co-author analyses
Title	For reference only
Date published	For reference only
Topics or Keywords (if available)	These are included with all articles listed in WoS, but not with Google Scholar records. These help with interpretation.

Chapter 4: Findings

This study explored the application of social network analysis tools to a co-author network of researchers. Grounded in social capital theory, this research was designed to inform the work of stakeholders in multi- and interdisciplinary research communities. Social network analysis tools were used to quantitatively measure the relationships among and connections between a group of researchers who won National Science Foundation awards in the Cyberlearning (CL) program between 2009 and 2015. Publicly available databases and search engines provided the list of CL grantees as well as the bibliographic information used to discover co-author relationships among this population. Web pages and publicly available datamining tool (Deep Insights Anytime Anywhere: DIA2) were used to collect attribute data.

In keeping with the IRB approval for this study, all identifying information has been removed from the results. The purpose of this research was not to identify key individuals in a network, but to explore the approach of using publicly available data to reveal meaningful information about the structural social capital in multi- and interdisciplinary communities. In the next sections, a description of the data is followed by the findings of this research related to these research questions:

1. What underlying social capital structures can be determined about a group of researchers from bibliometric data and other publicly available existing data?
2. What are ways social network tools characterize the interdisciplinarity or cross-disciplinarity of co-author teams?

Description of the Data

This study applied social network analysis tools to publicly available data of about the Project Investigators and Co-Project Investigators (PI/Co-PIs) who received CL awards from the National Science Foundation's (NSF) CL Programs during the years 2009-2015. The focus of the study was the co-author relationships among this group of researchers with 230 awards for 411 PI/Co-PI's. The co-author relationships among these 191 researchers formed the network that was analyzed in this study. To determine the structural capital within this network, a 191x191 adjacency matrix was created from the list of co-author pairs. Numbers in the cells of the matrix represented the number of times each pair co-authored together (also known as a weighted edge list).

After the network relationships were calculated using the adjacency matrix, the resulting 191-node network graph revealed a large component of 60 researchers and two medium components with 18 and 12 nodes respectively. The remainder of actors in the network resided in two sets of 5 nodes, six sets of 4 nodes, seven sets of 3 nodes and 23 dyads. (See Figure 1) Note: Figure 1 includes all 411 researchers; the left margin shows single nodes representing the 220 researchers for whom no co-author instances with another CL researcher were found

Attribute Data

As part of the first phase of research, attribute data was collected for all grantees. This attribute data included places of employment (and academic department, where relevant), domain of last degree obtained, and current research interests. Of the 191 researchers who had at least one co-author collaboration with another CL researcher attribute data was added to the adjacency matrix data (of co-author relationships) to reveal more contextual information about 114 of the co-author relationships shown in the following network graphs.

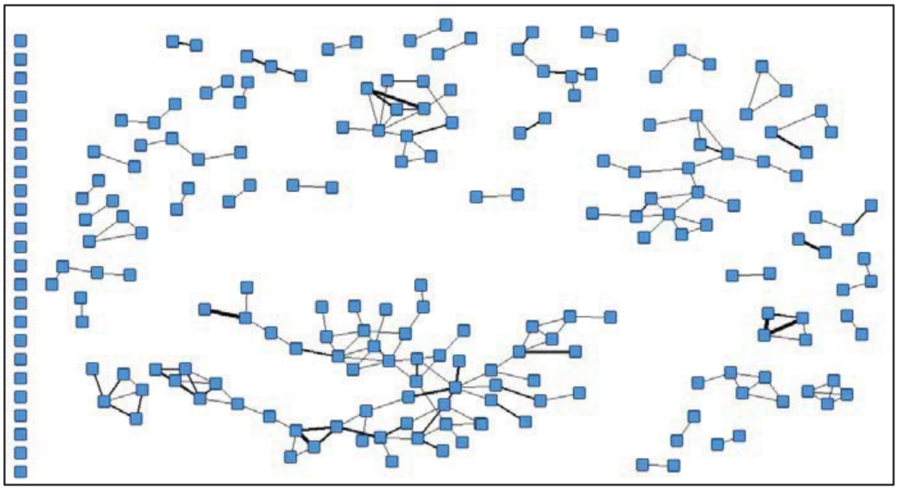


Figure 1. Network of 411 researchers

The following attribute data was included in the adjacency matrixes calculated within UCINet (Borgatti et al., 2002) calculations: Award Number, Domain, Number of co-authorship instances. In the resulting network graphs, these attributes are shown through node shape and color and the width of ties connecting nodes. The color of a node always represents researcher domain. Shapes assigned to each node based designate award number. The number of co-author collaborations is shown both through the thickness of the line (tie strength) and, in some cases, by a number that indicates the number of publications co-authored by each pair.

Attribute data takes on a larger role in a network of this size and density, where the data from the centrality calculations are not particularly illuminating. Additionally, a goal of interdisciplinary programs is often the collaboration of researchers across disciplines; the addition of domain attributes to the network data allows cross-domain collaborations to clearly and easily be identified. Relationships between nodes are quantified by the thickness of the tie between nodes, which indicates the number of co-author collaborations between two researchers. Further discussion of these attributes continues with Research Question 2, below.

Research Question 1

The first research question asked: What underlying social capital structures can be determined about a group of researchers from bibliometric data and other publicly available existing data? The following measures were run on the network of 411 nodes, of which 220 were solitary, unconnected nodes and 191 were connected to at least one other node though a co-author collaboration: Degree Centrality, and Eigenvector Centrality, Ego Homophily, Cohesion/Homophily, and Constraint/Structural Holes. Graphs depicting the relationships among 191 researchers in this co-author network are sparse and fragmented. Although Closeness centrality is recommended for information networks (Borgatti, 2005), Closeness centrality calculations don't produce meaningful numbers unless all nodes in a network are connected in a single component (Borgatti et al., 2013). Therefore, Closeness centrality was not included in this study.

Degree Centrality in an undirected network graph simply counts the number of ties that ego has to other actors in the network (Borgatti et al., 2013; Scott, 2013). The mean degree centrality for the network was 0.93. Figure 2 shows that those members of the network with the highest degree centrality typically reside in the learning science domain, except for a single human computer interaction researcher. It is interesting, therefore, to notice the outliers were two nodes in the main component of the network: A learning sciences researcher at a research institute and a human computer interaction researcher employed at a university, who each had a degree centrality of seven; this is of interest because of the high number of researchers categorized into Learning Science and Computer Science domains. Eight other well-connected nodes stood out for having a centrality between five and seven. Three of these 8 high degree centrality nodes are members of the same research institution (and part of the main component), which gives them an advantage of proximity; the remaining five were affiliated with different universities across the country

The remainder of the network explains the high standard deviation of 1.34; 81 nodes had a degree centrality between two and four, and 90 nodes had a degree centrality of one. Most the nodes with a degree centrality between 5-7 are found in the main component, where greater opportunities for connection exist; the 12 and 18 node components held one node each with a degree centrality between 5-7. (Of course the dyads were only able to reach a degree centrality of 1, and the 220 unconnected nodes had a degree centrality of zero.)

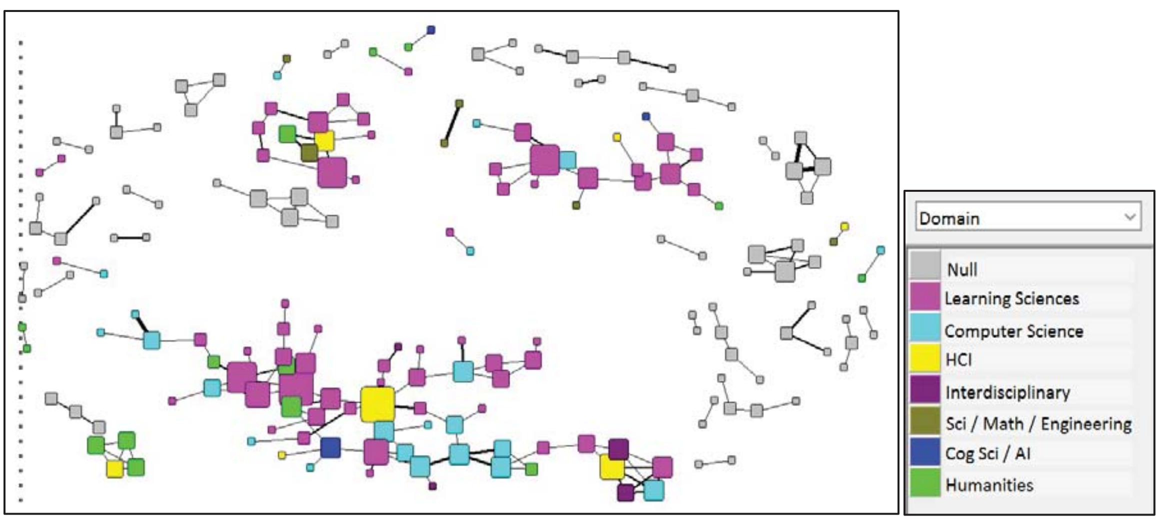


Figure 2. Degree centrality by domain

Eigenvector centrality is a more nuanced measure (than degree centrality) to measure the level of a node's connectedness within a network. A high Eigenvector number indicates the potential influence of nodes within the network; nodes with high Eigenvectors are naturally found in the larger clusters of a network, since those hold the greatest opportunities for connection with others. In small network, many of the same nodes that displayed high degree centrality numbers show up again with the top Eigenvector numbers. (Figure 3 shows only the main component of the network, because this was the only area of the network with notable variances in node size when Eigenvector centrality was displayed.) Both Eigenvector centrality and degree centrality provide an indication of which researchers in this co-author network are well-connected with other researchers. The inclusion of research domain as a node attribute in the network graphs shows whether co-author teams came from the same domain or if the team is cross-disciplinary.

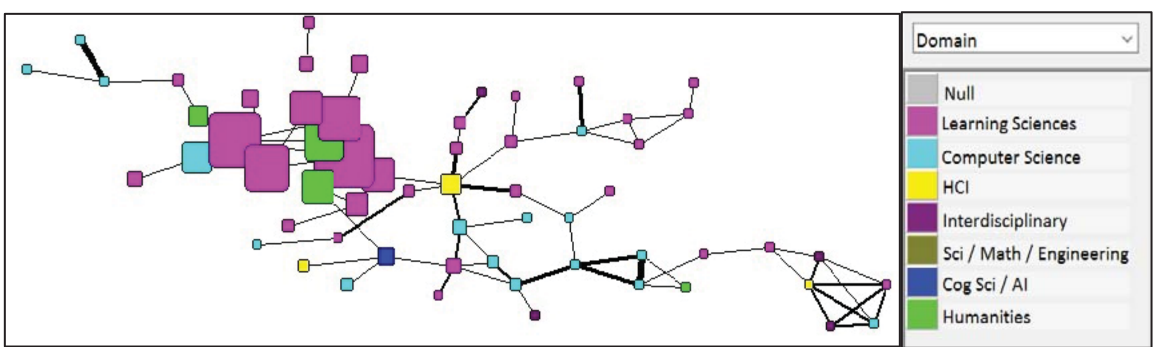


Figure 3. Eigenvector centrality in network main component

In this analysis, homophily was measured by calculating the network cohesion of this group with the E-I Index, which produced an E-I score for the network. The network cohesion analysis disregarded all marked with a zero; only for those 114 nodes for which domain attribute data had been added to the adjacency matrix data were considered in the homophily calculations. This network's E-I score was 0.0376 with a Correlation of 0.0237. This information shows a fairly balanced diversity among CL co-author teams, with researcher neither pairing exclusively with those from their own domains nor exclusively with those from other domains. UCINet's (Borgatti et al., 2002) network method of calculating structural holes determined network constraint for the entire network. In this calculation, the constraint index ranges from zero to one (or greater in smaller networks) (Hanneman & Riddle, 2005). Constraint numbers for the 191 researchers in this network with co-author ties ranged from .22 to 1.85 (the remaining 220 researchers with no ties had null values in this calculation). The mean constraint index was .87, with a standard deviation of .25. Of the 191 researchers, 111 of them had perfectly constrained networks with scores of 1 (or higher) meaning there were no structural holes to bridge. The ten lowest constraint values ranged from .22 to .41. Of the ten least constrained researchers, eight were employed at universities and two at research institutes; these researchers work in the learning sciences (6), computer sciences (2), human computer interaction (1), and cognitive science/artificial intelligence (1). The graph in Figure 4 indicates those researchers with the 10 lowest constraint scores with circles; those two researchers with the lowest constraint scores are circled in black and the remaining eight are circled in red. These results point to structural holes within the network and to bridging opportunities among like as well as different research domains.

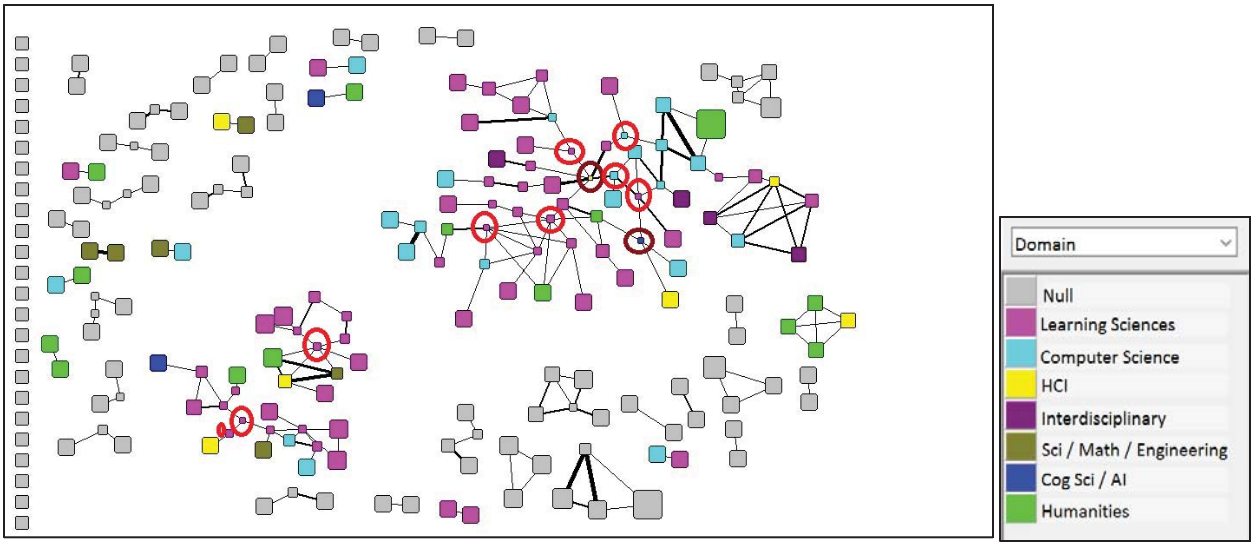


Figure 4. Structural holes shown by measures of constraint

After applying UCINet's (Borgatti et al., 2002) G&F Brokerage Roles function to the 191 connected nodes within the network, 37 nodes fell into at least one of the brokering roles. (See Chapter 3 for descriptions of these roles.) Table 4 indicates the number of nodes positioned in each brokering role, and the number of times each brokering role appears in the network (instances) of each brokering role appearing in the network. The co-authorship relationship is undirected, which makes leaves the Gatekeeper role (B→A→A) and the Representative (A→A→B) roles redundant; Table 4, therefore, combines these two roles into a single category. One final thing to note about the results of the brokerage role in this network: Attribute data was only collected for 114 of the 191 connected nodes in the network,[6] so these brokering roles apply only to this subset.

---

[6] See the section on Attribute Data in Chapter 2 for more details.

Table 4

Brokerage Roles and Results

Coordinator	**20 coordinators / 30 instances**
Gatekeeper-Representative	**13 gatekeeper- representatives/ 20 instances**
Consultant	**6 consultants / 6 instances**
Liaison	**9 Liaisons / 15 instances**

Of the seven domain categories in this study, two were not represented in any of the brokering roles: Humanities researchers and Interdisciplinary researchers. The 37 nodes with brokering roles in this network came from the Learning Science (25) nodes, Computer Science (7), Science/Engineering/Math (2), Cognitive Science / AI (2), and HCI (1). The distribution of these five domains across the brokering roles roughly mirrors the distribution of the domains throughout the network. The exception is humanities, which comprises 12% of the network population in this study and does not appear in any brokering role. (See Appendix D for a chart showing the anonymized 37 Nodes and their brokerages roles within the network.)

Research Question 2

The second research question is: What are ways social network tools characterize the interdisciplinarity or cross-disciplinarity of co-author teams?

The coding of the 114 researchers as a sample within this network gave some insight into cross-disciplinary nature of this co-author network. These 114 researchers were each assigned a to single research domain. Table 5 shows the distribution of CL researchers across the seven research domains assigned for this study. Learning Scientists comprised the majority of researchers in this sample. To answer the second research question, further attribute data was needed; award attribute data was added to the network data to meet this objective.

Table 5

Sample of Researcher Domains among CL Grantees

Researcher Domain	Number of Researchers	Percentage
Learning Sciences	58	51%
Computer Science	22	19%
Humanities	14	12%
Human Computer Interaction	7	6%
Science/Math/Engineering	6	5%
Interdisciplinary	4	4%
AI/Cog Sci	3	3%

UCINet's (Borgatti et al., 2002) clique function was applied to the network data to identify 4-member cliques. This calculation revealed five different cliques among the 191 co-author connections. Each clique represents a group of four researchers who co-authored at least one research paper together. It is possible (and likely) that a single paper united the 4-member clique. It is also possible that that various combinations of a clique's membership have published together beyond that single paper that brought together the clique. Among these four researchers in each clique, increased tie thicknesses indicates additional co-author relationships between clique members.

Cliques. The 191-co-author network contains thirty 3-node cliques and five 4-node cliques; no cliques were found with 5 or more nodes. This purpose of analyzing these five 4-node cliques was two-fold: To determine clique interdisciplinarity and award homogeneity within each clique. (Award homogeneity within a clique would reveal any PI/Co-PI relationships that exist within the cliques.) Figure 5 shows these five 4-node cliques out of the context of the greater 191 co-author network; this removes extraneous visual clutter allowing the reader more easily focus on the composition of the cliques. The nodes in the group of five 4-node cliques in Figure 5 represent nine different awards and researchers from six of the seven domains identified in this study.

In Figure 5, the nine different awards are represented by nine different shapes. In the case that a researcher received more than one award, the award number was chosen that linked the researcher to others in the clique. The node colors assigned in Figure 5 represent the same domain color codes shown in all previous graphs in this study. (Note: The nodes in Figure 5 consist of six of the seven possible domain color codes in this study.) Nodes with identical shapes indicate a group of researchers who were part of the same award team. Each of the five cliques is labeled with numbers 1-5. Readers notice that two cliques are represented within 4 different graphs, because cliques #4 and #5 are found within the 5-node graph. The following paragraphs analyze the 5 cliques.

In Clique 1, the lack the lack of common node shapes indicates that none of the members of Clique 1 participated on the same grant. The domain representation of this group is slightly mixed, with a one Humanities researcher among the three Learning Scientists. One single paper tied this group together.

Clique 2 shows is an example of a co-author group with maximally multidisciplinary membership; as indicated by the four different node colors each member of this clique belongs to a different research domain. Three researchers in this clique participated on the same award, which shows the cross-disciplinarity of this award team. Two pairs in Clique 2 have a strong co-authorship connection, but it is a single citation that connects the four. Clique 2 belongs to the component with 12 nodes and can be seen within the context of the network in Figure 2.

Clique 3 shows only slight cross-disciplinarity among its researchers; three of the four co-authors were assigned to the Humanities domain, while the fourth belongs to the Computer Science domain. Clique 3 is the most isolated and homogenous of the five cliques in this network and stands unconnected to any other nodes within the network. All four researchers in Clique 3 participated on the same award, are tied together by two citations that were generated from that award.

Cliques 4 and 5 are interconnected; the two cliques in this cluster are labeled abcd and bcde in Figure 5. Two different grants are represented in these two cliques and each clique is maximally diverse insofar as domain is concerned, which reflects the cross-disciplinarity of the two award teams. Cliques 4 and 5 lie at the bottom edge of the main component. These two cliques are loosely attached to the main component of 60 researchers.

Examination of cliques using both domain attributes and award number attributes gives stakeholders a sense of the interdisciplinary work done by researchers. This data makes it easy to gauge the cross-disciplinary research of teams that are publishing together early in the lifespan of the Cyberlearning community.

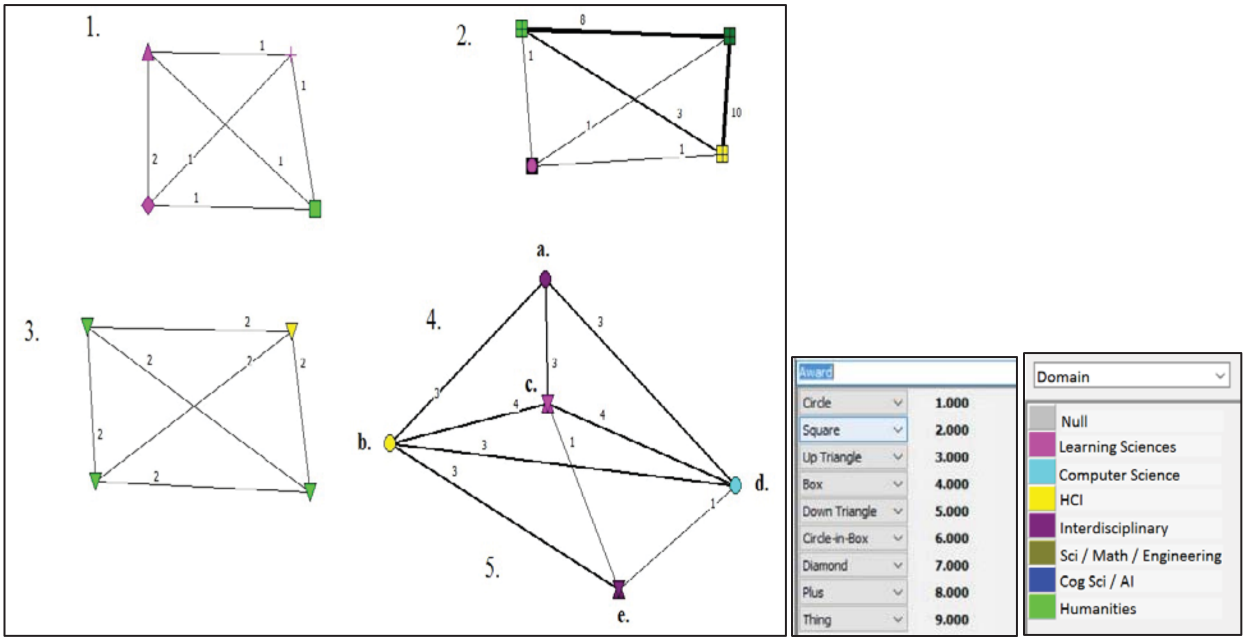


Figure 5. Cliques

Dyads. The network graph of 114 co-authors includes 23 sets of dyads. As the smallest unit of co-authorship dyads that include attributes for researcher domain and award number may be a simple way to identify at successful cross-disciplinary award teams. Ten dyads were selected at random from the 23 dyads in the network. Dyads were coded 0/1 depending on whether the co-author team also belonged to the same award team. The attribute of award team membership was expressed in node shape and the attribute of research domain was expressed through node color. In Figure 6, square-shaped nodes belonged to the same award team and circle-shaped nodes belonged to different award teams. The number between two nodes represents the number of publications produced by that dyad.

Eight of ten the dyads were made up of people who were part of the same award team. The two dyads that appear to be from different awards (because each co-author has a different award number) each belonged to dyad teams where the respective awards complimented one another; these two dyads also essentially come from the same award team. (The award numbers differed within these dyads, however the title of the awards was the same within the dyads.) Effectively, all ten dyad members were working on the same award. Five out of these eight dyads consisted of a cross-disciplinary teams (63%). The research domains of the Humanities and Science/Math/Engineering make up almost half of the membership of these ten dyads; this is noteworthy because these two groups make up a smaller percentage of the research network. Of these ten dyads, seven of them represent co-author teams that published a single paper. Two of the ten dyads published two papers related to their award. One dyad produced 11 papers, none of which appear to be related to their award. The ten dyads shown in Figure 6 are shown removed from the context of the network to make it easier to examine them; when seen within the context of the network, all dyads are isolated from other components and are placed at random on network graphs.

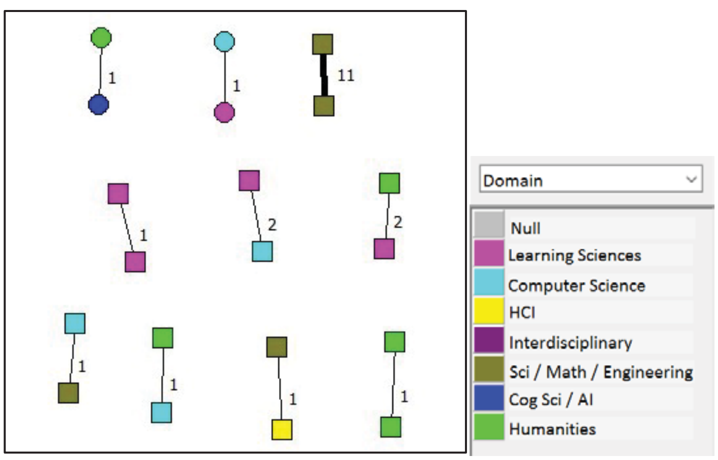


Figure 6. Ten random dyads

Conclusion

This research used social network analysis tools to calculate and graph elements of social structural capital that exists within a group of researchers by using co-author relationships as the unit of analysis. Of the 411 CL researchers, 191 of them co-authored with at least one other CL grantee; this network consists of one large component, two medium components, and several smaller clusters and dyads. Attribute data for these calculations included research domains and research award numbers. Degree centrality and Eigenvector centrality measures were used to calculate influence of researchers within the network. Constraint measures showed those researchers who occupy structural holes within the network. Due to the small size and fragmented nature of this nascent network, it was found that often those researchers who held positions of influence within the centrality measures were those who also occupied structural holes within the network. Homophily measurements calculate how often researchers were connected to other researchers from the same domain; these results indicate a fairly even distribution of connections among researchers. Calculations revealed five 4-member cliques with the network. These cliques and ten randomly chosen dyads were examined to learn more about the multi-disciplinarity among co-authorships by looking at the domain attributes assigned to each researcher. Multi-disciplinarity varied among the 4-member cliques, though no cliques were completely homogenous. Most dyads consisted of members from the same award team, and a little over half of these were multi-disciplinary teams. The most common brokering position (coordinator) is a homophilous role, with 20 researchers from the network in this role positioned between two nodes from the same domain. The second most common brokering role, with 13 researchers in this role, is gatekeeper/representative, gatekeepers are positioned to connect a researcher from their domain with a researcher from another domain. The most diverse brokering role is that of liaison; in this grouping the three nodes all hail from different research domains. There are nine researchers with a liaison role. (It is possible for one researcher to be positioned in more than one brokering role and for a researcher to have multiple instances of a single brokering role.) By using publicly available data, this method offers a glimpse into the social structural capital among a group of researchers. Analysis of these results follows in Chapter 5.

Chapter 5: Discussion and Future Directions

The purpose of this study was to explore what can be learned about the collaboration patterns within a group of cross- and interdisciplinary researchers through the application of social network analysis tools to co-author and other publicly available, existing data. This research is of potential interest to funders of multi- and interdisciplinary research and those centers, institutes, consortia, or organizations that exist to support the multi and interdisciplinary work of their grantees and to understand the growth, impact, and diffusion of ideas and research within their research communities. One of the ways that fields develop and how ideas flow through a community is via scholarly communication (Crane, 1972; McFadyen & Cannella, 2004; Wagner, 2008). In addition to acting as vehicles to spread ideas, scholarly publications in the forms of papers, chapters, books, conference proceedings and posters serve as evidence of researchers' collaborations. In that vein, this research used bibliographic citations of scholarly communication to identify co-author collaborations among researchers who received awards from the Cyberlearning (CL) program funded by the National Science Foundation (NSF). The choice was made to use public, existing data to avoid taxing extremely busy researchers with the requirement of a survey. The newness of the CL program, the potential for CL to develop into an actual field, and the interdisciplinary focus of the awards within the program presented an interesting case to explore the use of bibliographic and other publicly available data to gain one view of the social structural capital of a community.

Social capital is the theoretical framework supporting this research. Social capital is the real or potential value found in a network of social relationships (Bourdieu, 1983; Coleman, 1994; Putnam, 2001). Among a group of researchers, social capital matters because the flow of information through a scientific network is one of the ways that science is advanced (Crane, 1972; De Solla Price & Beaver, 1966). Co-authorship is studied here, because the relationships are visible artifacts of collaboration and communication among researchers. Co-authorships provide stakeholders with evidence of collaboration among researchers and can allow stakeholders to intuit the flow of real or potential knowledge within a network (Borgatti et al., 2012). Social network analysis tools provide the means of making these connections among researchers visible. While this research establishes baseline measures about collaboration in the CL community through co-authorship, this method, in general, could also be of interest to other funding agencies, institutes, centers or consortia that are designed to support interdisciplinary research. The specific research questions in this study are:

- RQ1. What underlying social capital structures can be determined about a group of researchers from bibliometric data and other publicly available existing data?
- RQ2. What are ways social network tools characterize the interdisciplinarity or cross-disciplinarity of co-author teams?

The population of grantees was obtained from the NSF's Award Search Database. The NSF datamining tool Deep Insights Anytime Anywhere (DIA2) and internet search engines provided the qualitative information such as employment/department affiliation, domain of last degree, and NSF award numbers that became the attribute data used in this study. Thomson Reuters Web of Science Core Collection (WoS) and Google scholar were used to collect citations of scholarly work published by those names retrieved from Award Search. WoS and Google scholar searches were limited to citations published between the years of 2009-2015. The social network analysis software UCINet illuminated the co-authorship network structure of the community; attribute data added context to the structure. NetDraw software was used to visualize and clarify both the structure and domain attributes of this community. This chapter introduces the study, includes a review of the relevant literature, and discusses its major findings, limitations, and implications for practice.

Relevant Literature

All knowledge is socially constructed, and the evolution of scientific research is no different. Pre-dating a formal education system, 17^{th} century scholars sent writings of their work and discoveries back and forth to one another in what was later dubbed the invisible college (Crane, 1972; Wagner, 2008). While single authorship is common, and there are some solo researchers, the creation of new research depends on collaboration and communication with others (McFadyen & Cannella, 2004). Scholarly publications provide evidence of both the direction of knowledge being created by researchers and of collaborations and communication among the researchers themselves. Taking into consideration the busy schedules of high-level researchers, the ability to gain a window on the collaboration patterns of a group of researchers though existing data should be welcomed.

As science has advanced and the world has grown more complex, the problems that need to be tackled are increasingly unanswerable by a single department or domain. Interdisciplinary research is a natural response to thorny problems, but disciplinary structures in institutions and funding organizations often make it difficult to support (Ledford, 2015). For these reasons, those stakeholders in interdisciplinary research are likely to have a particularly keen interest in finding out who their researchers are and who's collaborating with whom. Co-authorship patterns, when viewed through the lens of social capital, can provide valuable information to those who are positioned to influence or direct the focus of an area of research that can help inform their funding allocation or programming decisions.

Social capital theory. Social capital can trace its influences back to Karl Marx's Theory of Capital, which was both an economic treatise and well as a social critique. While social capital is concerned with the real or potential value that people find in their interpersonal relationships (Bourdieu, 1986; Burt, 2000; Lin 1999, 2001b), influences of neoclassical economics and sociology remain strong forces in social capital theory. Social capital, in contrast to human capital or economic capital, exists within a person's network, between individuals, rather than with the individual (Bourdieu, 1986; Burt, 2000; Lin 1999, 2001b). The structure of social relationships impacts the flow of information and the way knowledge is spread through a group (Burt, 1992; Lin 1999, 2001b); and the influence held by individuals (Bourdieu, 1986; Lin, 1999). In the case of looking at interdisciplinary research networks, particularly those that are newly formed, social capital is a useful framework.

One of the major themes of social capital theory is the concept of network denseness, which shows the relative cohesion of members of a network. A network where there are many connections among members of a group is dense, or can be said to have many ties. Some of the benefits of a dense network are trust, safety, and strong social support Bourdieu (1986), Coleman (1988), and Putnam (2001). More sparsely connected networks can foster innovation and entrepreneurship (Burt, 1992; 2000; 2004; Granovetter, 1973). This network holds potential for both. Though it is tempting to see these as diametrically opposing views or hold one model of density in preference to another, the benefits (and downsides) of network density depend largely on the context of the network itself. Coleman (1988) gave an example of the strong bonds of trust between New York diamond traders that allowed expensive gems to be loaned among businesses that would not have been possible, if members didn't fear any impropriety would have ramifications on their livelihoods. On the other hand, Portes (1998) discusses how strong ties can be stifling and inhibit innovation because members feel pressure to follow group norms.

Granovetter's (1973) work showed the benefits of an individual's loose ties and how those personal connections outside one's inner-circle can prove to be the most valuable. Burt (1992; 2000; 2004) extolled the virtues of structural holes, elements that exist in less-dense or between loosely connected groups within a network. In both instances, the value is on bringing together diversity either in individuals or groups. A person who is poised to connect two diverse nodes or groups is poised to either bring these two diverse nodes together or keep them apart; thus, the position potentially controls the flow of information between the groups or individuals. In the both network structure and social capital theory, such a person is called a broker. Traditionally, the broker role has been one in which the benefit to the broker lies in controlling access between two groups or individuals (Burt, 1992, 2000; Gould & Fernandez, 1989). However, there are other possibilities for those in brokering roles. Obstfeld (2005) illustrated a model where the role of the broker was to bring groups together. The coming together of new groups could lead to innovations. A case study by Lingo and O'Mahoney (2010) made clear that those who sit at the intersection of structural holes in brokering roles use a variety of techniques to both keep groups apart and bring them together, based on what circumstances require. Finally, it should be noted that the brokering role can come with constraints, for example, when those in that position feel pulled in different directions or constrained by having to follow norms in multiple groups (Krackhardt, 1999). Stakeholders must consider the context of their researchers' networks, before drawing specific conclusions about what these brokering roles might mean.

It is possible for those in positions to use the data from a social network analysis influence a network by creating connections where previously there were none or bringing together (Cross et al., 2002). In this case, researchers might be brought together though programmatic or funding means. A word of caution for stakeholders: The creation of new ties and maintaining existing connections comes with a cost (MacFadyen & Cannella, 2004). It can be assumed that the creation of new ties across a community come at a cost of some other. Indeed, one of the downsides of participating in interdisciplinary research can be the erosion of some ties within a domain-specific community (Ledford, 2015). It is important to remember that these ideas are theory, and not a guaranteed outcome.

Social network analysis. If social capital, colloquially, can be defined simply as who knows whom, then the structure of the relationships can be understood as an indicator of how knowledge flows through a network (Burt, 1992). Social network analysis is a mathematical way to understand these relationships and perform calculations that highlight certain positions within a network. The mathematical relationships can be graphed so that relationships in a community can be visualized. Social network analysis can be used both to explain past experiences, project forward what might happen, or provide information to plan future work. The ability to recognize the actors in a co-author network who are well connected to others like them or who bridge across structural holes is useful information for stakeholders in interdisciplinary research communities. Structural cultural network graphs and data can also inform stakeholders of researchers who are well connected with others in their discipline, or those who write with cross-disciplinary colleagues. Information learned from applied social network analysis techniques makes networks visible and provides descriptive information that can be used to inform decision making or actions (Borgatti et al., 2013).

A traditional social network analysis ideally includes a survey that includes all members of a group or organization to get an accurate picture of a network. To avoid taxing an already busy group of researchers with a survey that would likely yield a poor response, this study relied on existing public data. An alternative to surveying a community is to apply social network analysis techniques to existing data that represents communication and collaboration (Borgatti et al., 2013; Long et al., 2013). This research used publicly available existing data in the form of the co-author relationships found in bibliographic citations of scholarly publications.

Bibliometrics and scientometrics. In 1960, Eugene Garfield founded the Institute for Scientific Information and began the Science Citation Index in 1961 out of his insight into the importance of bibliometric data in understanding the growth of scientific knowledge (Garfield, 1979). Newman (2004) observes "the co-authorship network is as much a network depicting academic society as it is a network depicting the structure of our knowledge" (p. 5200). The term *bibliometrics* is older than scientometrics and measures the output of scientific literature in the form of author influence, citation counts, co-citation analyses and number of publications (Ding, 2011). Scientometrics uses publications to measure impact for policy makers and funding agencies (Leydesdorff & Milojevič, 2015; Van Raan, 2005) and uses often highly sophisticated visualization tools to create domain maps of scientific research (Boyack et al., 2005). Both fields are concerned with growth of science and the establishment of disciplinary boundaries (Godin, 2006; Leydesdorff & Milojevič, 2015; Van Raan, 2005). Thomson Reuters' Web of Science is the modern version of Garfield's Science Citation Index. SCOPUS is a relative newcomer and Web of Science's only true competitor. These two sources of bibliographic citation information are both expensive products that justify their costs by providing indexing and the ability to download and save data to multiple formats to support different types of bibliographic research.

Types of analyses used. The UCINet calculations employed in this research were: Degree Centrality, Eigenvector Centrality, Homophily, Constraint, and Brokerage. Two structural components analyzed in this research were cliques and dyads. An overview of the literature pertaining to these analyses follows.

Degree centrality is measured simply by the number of direct connections each node has to other nodes in a network (Borgatti et al., 2013; Scott, 2013). Eigenvector centrality is like degree centrality in that nodes are counted, but this measure also weights all nodes according to their relative connectedness. Thus, Eigenvector centrality can be understood as a node's propensity to be exposed to whatever is traveling through a network (Borgatti et al., 2013). High centrality nodes in this network are well positioned to receive and pass on news and new ideas, or even to bring various people together.

Homophily is a measure of sameness in a network. This study looked at domain attributes to learn how often researchers were publishing with others from the same domain and how often co-author teams consisted of researchers from differing domains. Constraint is a social network analysis measure that can be applied to a network to identify structural holes.

Constraint, in simple terms, is a measure of how many of a node's ties are connected to one another; the more connected a node's network, the more constrained it is (Borgatti et al., 2013; Hanneman & Riddle, 2005). The idea of structural holes stems from social capital theory. Structural holes are breaks in or thin connections between clusters within a network; those network members who can span structural holes are positioned to either benefit from information or pass on information (Burt, 1992; 2000; 2004). Structural holes can neither be classified as good or bad, whether they are useful or a situation to be remedied depends entirely on the context of the network. Coleman (1988) and Putnam (2001), for example, take the view that a tightly knit network with few structural holes is a desirable for its ability to create trust within a community.

As opposed to Burt's neoclassical economics view of structural holes, Gould and Fernandez (1989) talk about the ability of a node sitting between two others to play a more altruistic, or at least non-competitive brokering role among and between groups. Furthermore, in their 1989 paper, Gould and Fernandez developed 5 different types of brokering roles, depending on the composition of the two parties being brokered: Coordinator, Gatekeeper, Representative, Consultant, Liaison. Because multi- and interdisciplinary teams are of interest to this study, it's worth noting that the Coordinator role is homophilous. Additionally, in this non-directed network, the roles of Gatekeeper (A→A→B) and the Representative (A→B→B) are effectively the same; both roles have two same domain nodes with one of those two occupying the central position. It is worth noting that only those nodes connected to two others meet the baseline conditions for brokering; any unconnected nodes and dyads could not be included in the brokering calculations.

There are three points to understand about cliques: These network subgroups are composed of a group of nodes that are all equally interconnected; the smallest clique size is three nodes; cliques overlap is possible (Borgatti et al., 2013; Scott, 2013). Dyads simply represent two researchers who co-authored at least one publication together and who have no ties to any other components in the network.

Methodology

There are three ways of constructing a co- author network: Drawing from a known population of researchers (Börner et al., 2005; De Solla Price & Beaver, 1966; Milesi et al., 2014; Rodriguez & Pepe, 2008; Yin et al., 2006); creating the co-author network from citations found searching by keyword or journal in citation indexes (Barabâsi et al., 2002; Bordons et al., 2015; Glänzel & Schubert, 2005; Li, Liao & Yen, 2013; Liu et al., 2005; Newman, 2004;); or beginning with a set of research papers (Xian & Madhavan, 2014). For this study, the first method for constructing the co-author network was used and a list of CL grantees was what set the boundary. Whatever sets the boundary on the co-authors should be the best fit for an organization, institution, agency, or center doing the research.

Six publicly available sources provided the data for this research: NSF Award Search, Deep Insights Anytime Anywhere (DIA2), Thomson Reuters Web of Science Core Collection (WoS), Google Scholar, and either organization/department websites or personal websites. Data collection was a two-part process. First CL awards and researcher names were downloaded from NSF award search and attribute data for researches was obtained from DIA2 and search engine results. Second, citations of researcher publications were downloaded from WoS and Google Scholar. Data processing and analysis required the collected data to be cleaned, sorted, and formatted for UCINet. One of the attributes for the researchers was domain of research. Domain categories were coded using the final degree earned, current research domain, and research interest(s), where available. UCINet and its accompanying visualization software NetDraw were used to calculate appropriate centrality measures and to identify positional roles within the network. As described above, the types of analyses used were: Degree Centrality, Eigenvector Centrality, Homophily, Constraint, and Brokering. Results are discussed below.

Sample

Four hundred and eleven PI /Co-PI names were extracted from=the Award Search download of 230 grants with the 8020 CL award code. A name search conducted in WoS and award search in Google Scholar resulted in 4006 publication citations. A list of author and co-author names was reduced to only those groups of CL researchers who co-authored with one another. Of the 411 CL PI/Co-PIs, 191 of them belonged to the network of co-authors.

Results

With only 191 of the 411 researchers connected through co-authorships, as anticipated, there was a lack of density. A NetDraw generated graph of the entire network revealed a large component of 60 researchers and two larger components with 12 and 8 nodes respectively. The rest of the network was made up of two sets of 5 nodes, six sets of 4 nodes, seven sets of 3 nodes, and 23 dyads (See Figure 7). The 220 grantees who were not connected to another grantee through a co-authorship are not included in Figure 7; these unconnected nodes add little to the visual and are shown as a line of unconnected nodes along the side of the graph.

(See Figure 1 for the 411-node graph.) The only attribute data included in Figure 7 is the tie strength between two nodes. Tie strength represents the number of times two authors have co-published articles together; the thicker the line between two nodes, the more often that pair has published together. Though Figure 7 captures an early snapshot of co-author collaborations among a recently-funded research area, it is important to realize some grantees may share a history of collaboration that predates the CL awards.

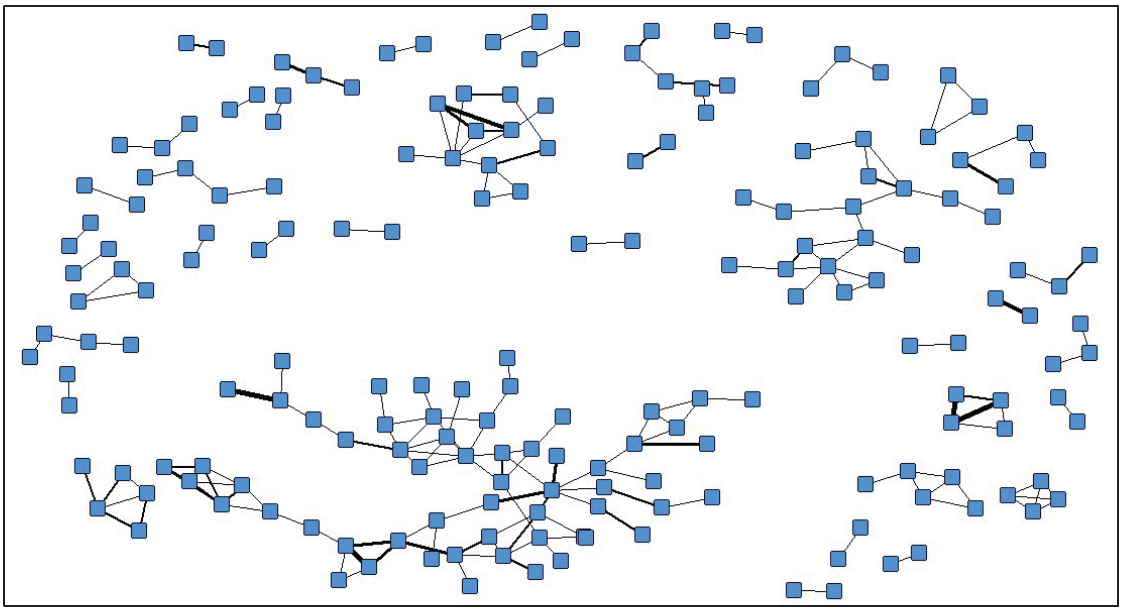


Figure 7. Network of 191 co-authors

Degree centrality results. Across the network, those with the highest degree centralities (between 5-7 degrees) came from the Learning Sciences and Human Computer Interaction. (See Figure 8) Researchers from the Learning Sciences comprise the majority of this sample (51%) (See Table 5); it may therefore not be a surprise that members from this group occupy some of the most central positions within the network. Human Computer Interaction researchers make up only 6% of the sample. At a degree centrality of four, other domains entered the picture: Computer Science (six researchers who had a degree centrality score of four) and one each from four other domains. This shows that many domains are represented as central players within this co-author network, which may be welcome news to stakeholders who have an interest in cross-disciplinary collaborations. The unexpected appearance of certain groups in central locations within a network may provide stakeholders with insights to the network and may also serve as a foundation for more exploratory research.

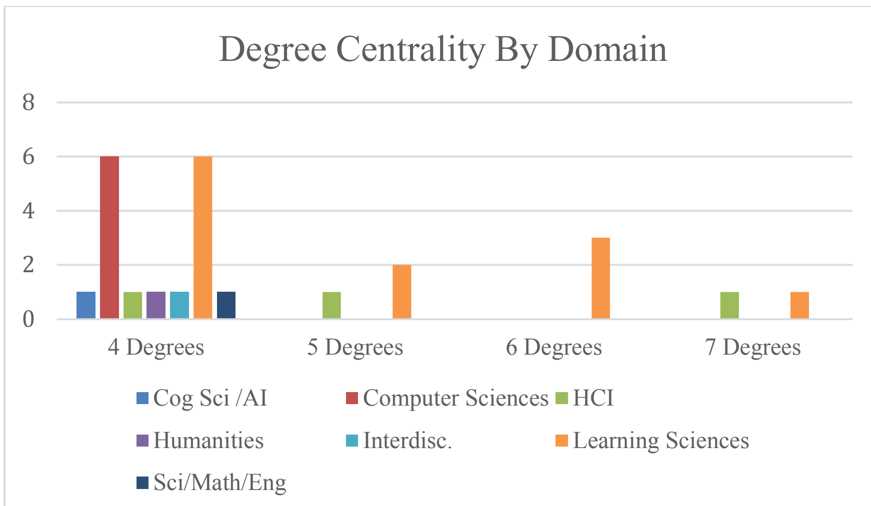


Figure 8. Degree centrality by domain

Eigenvector results. Half of those researchers with high Eigenvector numbers work at a research institution and the other half work at universities. While those who are affiliated with a research institution make up only 7.3% of the population, those who work at universities make up 83.2% of the network. This data shows that those affiliated with research institutions are highly connected with others, who are also highly connected within the network – at a disproportionately high rate. Figure 9 shows Eigenvector calculations from a detail of the network; the cluster of nodes with high Eigenvector number are largely from Research Institutions.

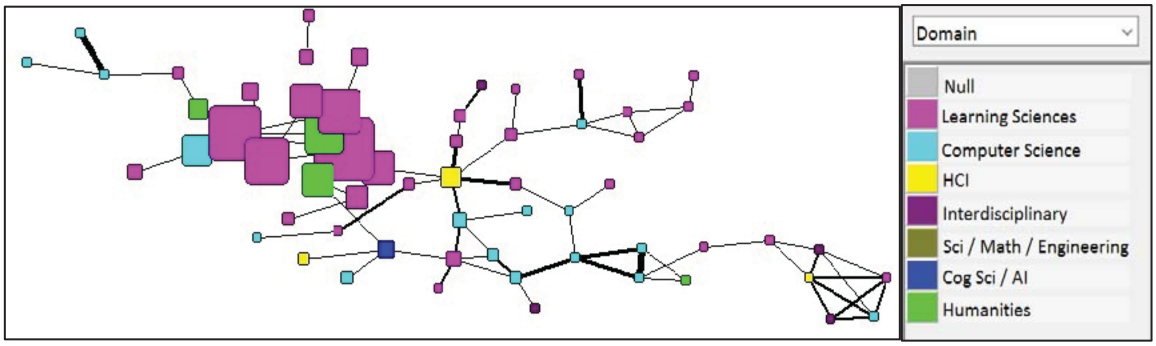


Figure 9. Eigenvector centrality in network main component

Homophily results. The Learning Sciences and Computer Sciences researchers, who make up over 70% of the network, overwhelmingly published with those from their respective domains. However, these groups also published with researchers across all six other domain categories in this study. The Human Computer Interaction and Humanities were the next diverse publishing groups; each of these groups published with 4 domains other than their own. Homophily was measured across the entire network. The final E-I score given for the network was calculated by looking at the ties of each node in the network, where domain information was available, the number of ties to researchers in the same domain and in each other domain was calculated and that returned a single E-I score of 0.0376. This E-I score indicates that the network was fairly balanced among those researchers who had co-authors from the same domain and those who belonged to multi-disciplinary co-author teams – tipping slightly towards more diverse teams.

Constraint results. All major clusters in this network have at least one member with a low constraint score, these nodes are essentially the glue that holds the individual components together. Of the top 10 least constrained network members (i.e., those occupying structural holes) the top two belong to the Human Computer Interaction Domain and the Cognitive Sciences / AI domains. (See Figure 10) Of the remaining eight nodes, six of them belong in Learning Sciences and two in Computer Science. Learning Sciences and Computer Science combined make up 70% of the population. Researchers affiliated with universities occupy the top six least constrained positions; two members from a research institution are also in the top ten least constrained nodes.

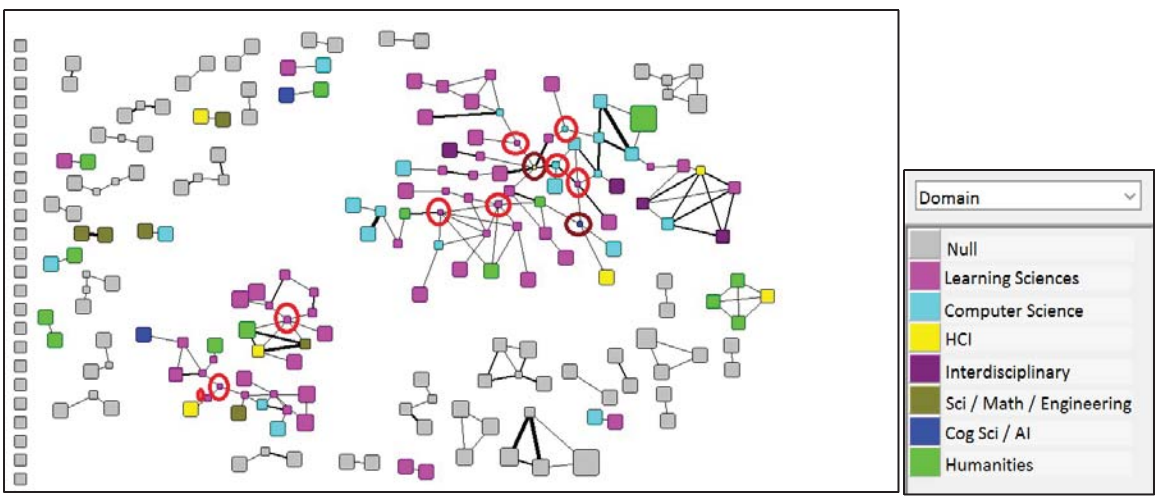


Figure 10. Structural holes shown by measures of constraint

Brokerage results. When the application was applied to the network, 37 nodes were found to be in one of Gould and Fernandez's five brokering positons. Of the seven domain categories in this study, the learning sciences accounted for 68% (25) of the brokering roles. The Learning Sciences occupied the most brokerage positions, three members of that domain occupy five different brokering roles. (See Appendix E for examples of the five brokering positions within the network.) The most occupied brokering in the network is that of Coordinator (A→A→A), with 20 different nodes in that role with 30 different brokering opportunities, however, since this is the homophilous role and one interest of this study is to look for cross-disciplinary co-author collaborations, the Coordinator role is less interesting. Consultant (A→B→A) was the least occupied brokering role in the network with only 6 nodes, each with one instance of being a consultant. The most heterogenous brokering role is Liaison (A→B→C) found nine nodes occupying that role with 15 potential brokering routes. The combination of Gatekeeper (A→B→B) and Representative (A→A→B) yielded 14 brokering positions and 20 potential brokering opportunities. Those in the Learning Sciences domain garnered 52 of the 71 potential brokering opportunities. The Humanities domain did not have a single node occupying a brokering role. (See Appendix E for graphs highlighting nodes positioned in the brokering roles.)

Clique results. In this study, node size was set to 4 nodes and five cliques were found. (See Figure 11) Four out of five of the cliques were part of larger network component, while one sat in isolation. Researchers in the isolate clique co-authored two publications together; all clique members work at the same university and were members of an award team. This study revealed several noteworthy clique and clique membership. One of the cliques, Clique 2, had strong author ties among nodes from diverse domains. Three of the four were members of the same award team and the clique was connected via a single publication. All Clique 2 members work in higher education and three of them belong to the same award team; in this case, a clique represents part of a successful award team, in that they published together. Clique 2 sits at the center of the 12 node component, and one of the members has a high degree centrality, low constraint, and a high brokerage number.

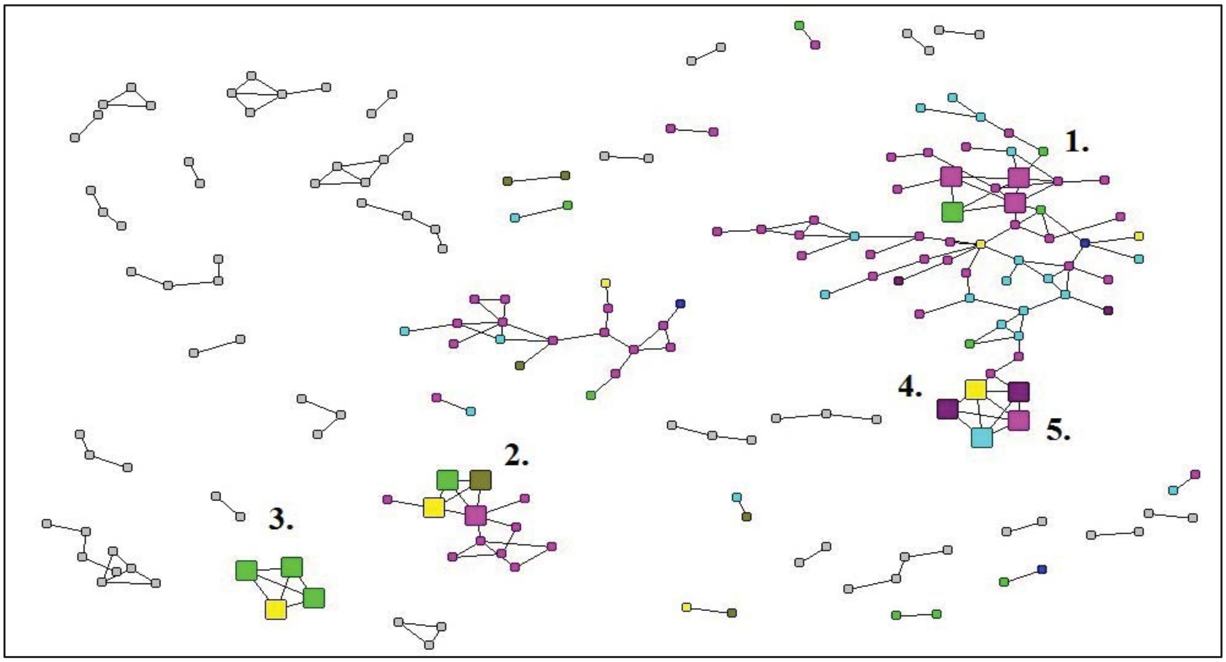


Figure 11. Cliques within the larger network

Dyad results. Of the 23 dyads in the network, ten randomly selected dyads with were examined to see what could be learned about the interdisciplinarity of research teams in this network. This sample revealed 7/10 of the dyads formed multi-disciplinary co-author teams. Among the ten dyads, six of the seven domains assigned to this study were represented: Learning Sciences (5 nodes), Humanities (5 nodes), Computer Sciences (4 nodes), and Engineering (4 nodes). Award numbers were included in the dyad attribute data. Dyads in which both researchers shared the same award number were assigned square node shapes and dyads in which researchers had differing award numbers were assigned circle node shapes. When the attribute data was assigned to a graph, two of the ten dyads were assigned circular nodes. In all cases the dyad members worked on the same grant; in two instances researchers had collaborative awards[7]. The tie strength between dyads represents the number of articles published by each dyad. (See Figure 6 for a graph of the ten random dyads.*)*. In all 22 articles were published by the 10 dyads. The strongest co-authorship tie among these dyads stems from two Engineering researchers, whose 11 co-publishing collaborations predate their 2015 CL award. Table 6 shows the researcher domains of dyad members and the number of publications produced by this team during 2009-2015.

Table 6

10 Random Dyads with Domains and Publications

Name	Domain Partner A	Domain Partner B	No. of Publications
Dyad A	Humanities	Cog Sci/AI	1
Dyad B	Computer Science	Learning Sciences	1
Dyad C	Sci/Math/Engineering	Sci/Math/Engineering	11
Dyad D	Learning Sciences	Learning Sciences	1
Dyad E	Learning Sciences	Computer Science	2
Dyad F	Humanities	Learning Sciences	2
Dyad G	Computer Science	Sci/Math/Engineering	1
Dyad H	Humanities	Computer Science	2
Dyad I	Sci/Math/Engineering	HCI	1
Dyad J	Humanities	Humanities	1

Discussion

Research question one asked: What underlying social capital structures can be determined about a group of researchers from bibliometric data and other publicly available existing data? UCINet was used to calculate basic centrality measures that identify which of the

[7] Collaborative Award dyad members received unique award numbers, but shared the same grant title; in these instances it appears two researchers from two different grant published together, however these dyads are part of the same grant.

191 researchers in the network lie are centrally located in the potential communication paths in the network of 411 potentially-connected researchers. This study examined the homogeneity of the network and ran two measures that stem from social capital research: Structural holes and brokering. Discussion of these results follows.

Degree centrality and Eigenvector centrality discussion. Borgatti et al. (2013) stated that most networks are not particularly dense. It was anticipated by the work of Börner et al. (2005) and Guan and Liu (2014) that the CL co-authorship network would reveal itself to be highly fragmented. The size and fragmented nature of this network contributes to many of the same nodes showing up in both degree centrality and eigenvector centrality measures. Eigenvector centrality analysis revealed two Humanities researchers who were highly central, although the Learning Science researchers dominated most centrality measures. It should be noted that the two researchers categorized into the Humanities domain work primarily in teacher education. However, both have interdisciplinary backgrounds and might just have easily been classified in the Learning Sciences. Indeed, most of the Learning Sciences researchers who occupy central positions in the network are interdisciplinary in their own right. Five of the top 10 researchers with the greatest Eigenvector centrality in this network are employed at the same research institution. In a 2005 study by Börner et al., it was seen that there are different co-author patterns depending on whether a researcher was employed at a research institution or a university; those at research institutions had higher levels of collaboration than those at universities. This same pattern appears to play out among these centrality measures.

Homophily discussion. E-I scores range from -1 to 1. An E-I score of -1 would indicate complete homophily; in this case, an E-I of -1 would mean that all network members published only with those in the same domain. An E-I score of 1 indicates complete heterophily; in this case, an E-I score of one would mean that all researchers in the network published only with those outside of their own disciplines. An E-I score of zero would indicate complete balance, With an E-I score of 0.0376, this network's publication patterns indicate a near balance between same-discipline and cross-discipline collaborations among those who are publishing. One of the expectations of the CL program was that all research groups would contain interdisciplinary expertise (NSF, 2011; 2014). Given that the distribution of domains is not evenly distributed across the network, 0.0376 is a favorable E-I score. This may be welcome news to stakeholders in this group who value multi-disciplinary collaborations, especially given the difficulties in communicating across domains and considering the relative youth of this network. While the results here don't show an overwhelming tendency to publish across disciplines, given the patterns seen here, a continuation of the trend toward heterophily seems likely.

Constraint discussion. Constraint is a measure that reveals structural holes within a network. Structural holes, an important concept in social capital theory, are important to identify. According to Burt (1992), those who span structural holes are poised to influence the flow of information through a community and potentially have access to diverse ideas and points of view. When domain attribute date is applied to a network, stakeholders in interdisciplinary work can see who pivotal players are, what groups are not communicating, and who might benefit from being connected. Being able to view structural holes in a research community with domain attributes could help guide stakeholder decision-making or policy implementation.

A lower constraint number indicates less constraint and more connections with those not tied to one another. Sparsely connected nodes in a network receive a higher constraint number. At this early stage in the CL research network, identifying structural holes is helpful when the actor occupying that structural hole is well-connected to others from different domains, because one of the core values of this network is multi- or interdisciplinary work. In this network, many of the same nodes with a high degree centrality are the same nodes with low constraint numbers. Six out of ten of the most constrained researchers in this co-author network belonged in the learning sciences domain. The majority of researchers in this co-author network are from learning sciences. Stakeholders may wish to look at those researchers with low constraint numbers, whose domains make up a smaller percentage of the overall network. This could be particularly valuable, if there is a desire to increase the involvement of that domain within the network. In a co-author network, those who occupy structural holes have worked closely with those to whom they have ties.

In this network, Learning Science and Computer Science researchers make up 70% of the population. In an analysis such as this, knowing who the researchers are who are poised to bridge structural holes could be valuable to stakeholders to get a glimpse of the network's social structural capital that has been created through co-author collaborations. By including affiliation as an attribute, as was done in this research, it is possible to know which organizations are represented by those nodes occupying structural hole positions. When a network includes membership from multiple organizations, stakeholders may be interested to know which organizations are positioned to span structural holes, in addition to the domains in which the researchers work. This information can provide insights into partnering opportunities or connections that stakeholders may wish to nurture.

Brokerage discussion. It is possible for a single node to sit in more than one of the five brokering positions and to have multiple instances of brokering within a certain position. (See Table 5 to review the Brokerage positions.) For stakeholders in interdisciplinary or multi-disciplinary research, the details of which domains sit in various brokering positions may be less interesting than the ability to identify which researcher from one domain is connected to researchers in another domains. Additional attribute data would make the knowledge of those in brokering positions more useful to stakeholders. For example, if a taxonomy of different types of CL were used as an attribute for nodes, it would be possible to see which Augmented Reality node is poised to act as in the brokering role of Liaison between a Cognitive Tutoring node and an Embodied Learning node.

In the case of this study, the domains are distributed almost evenly throughout the network. This can be observed by looking at the network graphics that include domain attributes and can be proven by calculating the E-I score. This can be seen in brokering, too; none of the less represented domains occupy the Coordinating positions (A->A->A). In this network, there are not enough minority domains clustered together in ways that would allow them to fill such a role. Throughout this network, few Humanities scholars occupy brokering positions of any sort. This lack of Humanities scholars in key positions does not necessarily indicate marginalization of Humanities researchers in the CL community, it could be a result of Humanities researchers comprising only 12% of the network. (See Table 5 for distribution of CL researchers among the seven domain categories in this research.) Over- or under-representation of a domain in brokering roles may be something stakeholders may wish to explore.

It is important to remember that the brokering roles assigned to nodes in the network reflect their position within the co-author network and these roles represent potential action or influence rather than active, realized action or influence. Additionally, while this study uses social network analysis tools, it is not a traditional social network analysis, rather, social network analysis tools were applied to a co-author network. As such, a co-author network represents only a sliver of the social structural capital that exists within a research community. Nevertheless, brokerage positions show where single-domain and cross-domain publishing collaborations occur among a group of researchers and may be of interest to interdisciplinary research stakeholders and aid in decision making and strategic planning.

The second research question asked: What are ways social network tools characterize the interdisciplinarity or cross-disciplinarity of co-author teams? This study used a method of focusing on interesting points of activity within the network (rather than analyzing the entire fragmented network) suggested for use by Schlager et al. (2009). The network units analyzed for Research Question Two were five four-member cliques and ten randomly selected dyads. The interdisciplinarity or multi-disciplinarity of research teams located within these subgroups was identified by adding the following attribute data to the selected subgroups: Researcher domain, employer affiliation, and award numbers to the network data.

Cliques discussion. By adding the context to cliques both with attribute data and social network analysis tools, it is possible to find potentially influential nodes. While anyone familiar with a research community is likely aware of the work of those researchers with a pivotal network roles, the information available through this kind of analysis provides a network context not otherwise available and gives stakeholders a richer, more holistic (though not complete) picture of relationships and connections among that research community. Having data about successful teams of researchers, their affiliations, and research domains provides stakeholders with small samples of co-author teams that could be interviewed to learn more about the factors contributing to successful collaborations across geographical space. Stakeholders may wish to interview members of isolated cliques when doing program assessment, since an isolated co-author team made up entirely of a PI/Co-PI team represents a cohesive, successful team that may be able to provide insights into measuring the impact of an award.

Dyads discussion. Dyads provide examples of award teams whose members also published together. Nine of these dyads published articles relevant to their CL awards, for this reason, dyads are of interest to stakeholders. One outlier dyad represented two Engineering researchers who had co-published 11 papers together prior to obtaining a CL award; these two are an example of a successful co-author team that applied for an award together. In the early stages of a research network, it is likely that any nodes with strong co-author ties represent researchers with prior writing collaborations. In the early stages of a network, stakeholders interested in multi-disciplinary collaborations may wish to scan a co-author network for strong ties between nodes from different domains to identify any naturally occurring, strong, multi-disciplinary ties.

The Dyads are isolated from any other network components; their position on a network graph is random and does indicate how closely tied a researcher may be to the rest of the network. Were a second-time slice analysis of the network performed, these dyads might become connected to larger components or reveal themselves to be the founding members of a larger separate component.

Dyads may leave the impression that they are representative of the entire collaborative writing team. In fact, of the 22 articles represented by these ten dyads, only two of the papers were written exclusively by a dyad team, the remainder involved other co-authors, who did not win CL awards and are therefore not represented within this network. This is of course the case with dyads, the cliques listed above, and all co-author connections in this study: the credited co-author team may in fact be larger than is reflected by looking at grantees and the CL community of co-authors is certainly larger than is represented by calculating co-authorship connections among CL award winners. These dyads provide examples of award teams whose members published together. Other co-author team members for these papers may be graduate students or other members of the CL community, but who are not visible in this study because they did not win a CL award, therefore the research domains of co-authors outside the CL community are beyond the scope of this research.

One of the main points gleaned from this exploratory research is that a co-author network analysis can be used to identify cross-disciplinary collaboration among research teams and cross-disciplinary collaborations that occur in co-authorships throughout the network that may not be tied to awards. . While this method may provide indicators of interdisciplinary collaborations by making visible co-author teams from different domains, this type of analysis cannot measure interdisciplinarity. Interdisciplinary research is a synthesis of research methods, tools, data, and vocabulary from multiple domains (Committee on IDR, 2004); this method does not examine the methodology behind the collaborations or explore the proposals for signs of interdisciplinary work.

Limitations

There are several notable limitations to this research. First, the co-authorship numbers were rather low. Second, there are limits to using co-authorship as a method of collaboration.

Low co-authorship numbers. After searching the names of the 411 CL researchers' names in WoS database and searching the award numbers of 230 CL projects in Google Scholar, the list of authors and co-authors from the resulting citations was narrowed down to only those citations that included two or more CL researchers. The list of CL researchers who co-authored with one another contained 191 researchers – less than half of the original list. In order to explain this low number, the award years of the grants received by the 191 co-author team members were contrasted with those 220 CL researchers who did not show up as co-authors. It was not the case, as expected, that award year might determine the presence of a co-author relationship. Both the 191-member group of co-authors and the 220-member group of unconnected researchers were well-represented throughout the CL award timeframe. Award type[8] did also not appear to be an indicator of co-authorship, since all CL award types were represented by members of the 191-member co-author network.

The relatively low percentage could be due in part to the use of Thomson Reuter's WoS database to collect the data. WoS focuses on high impact journals, which could have limited

---

[8] See Appendix G for information about CL award types.

the number of co-author collaborations discovered. Without a doubt, Google Scholar would have returned a greater, more representative number of citations. In the design of this research use of Google Scholar was deliberately minimized, because advantages gained from greater search results were outweighed by known issues of using Google Scholar for bibliometric research. The primary problem with use of Google Scholar for this type of research is the inaccuracy of citations and non-standard citation formats that are common to the search engine (Harzing & Van der Wal, 2008). Additionally, Scholar is not well-designed for even small numbers (<200) of citation downloads, as batch downloading is not possible. These issues created insurmountable barriers to conducting author searches using the non-automated methods in this study. To take advantage of the bigger pool of citations found in Google Scholar this study restricted Scholar searches to the NSF award number. Anecdotal searches of Google hint that the number of actual co-author collaborations is higher than the WoS results indicate. It is anticipated that the number of co-author pairs would have been higher, had SCOPUS citation database also been used, in conjunction with an author search in Google Scholar.

Limits of grantee co-authorship as an indicator of collaboration. Looking at co-authors among the CL grantees can show something about the formal collaborations between members of that community, but cannot account for graduate students who may have also worked on these projects or other researchers who may have influenced these projects. Due to limited resources, the co-author network in this study was limited to the collaborations of CL grantees; the ties in this network don't represent any co-author collaborations with researchers who were not members of CL award teams. This means that any CL researcher or graduate student who was listed as a contributing author but was not named in a CL award was left out of this network. Further, co-authorships networks create a tiny view into collaborations and social ties among researchers, since researchers who speak to one another at conferences and communicate about problems virtually may be quite connected. This research method does not account for other types of collaborations or project or communications among the researchers in the community that would give a sense of who's working with whom (on what). This study relied on publicly available data and unfortunately such informal interactions rarely leave publicly available artifacts. Most importantly, as a measure of impact, this method in no way accounts for projects, programs, or conference sessions and presentations. Products, methods, and interpersonal connections that resulted from awards may be among the more important measures of impact and cannot be accounted for with this method.

In this study, bibliographic data from 2009-2015 was collected for all grantees who won CL award between 2011-2015. This study did not examine which co-author collaborations occurred before awards were granted, or after, or whether award teams published together both before and after winning a CL award. A research design that accounts for the publication date relevant to the award date would also provide stakeholders with insights about the strength of ties in a research community. Adding publication dates and award dates to a network's attribute data could provide stakeholders with an understanding of a network's growth as that area of research matures.

Methodological Implications

This exploratory, small-scale application of social network analysis tools to existing, publicly available data was done to gain insights into the structural social capital of an emerging research community. None of the methods employed here were new, in fact, with datamining techniques taking center stage for large-scale bibliographic analyses (Bordons et al., 2015; Börner et al.,2005; Xian & Madhavan, 2014), the methods employed in this study were artisanal; this may have allowed a closer look at the data sources and intricacies of the data collection and cleaning process than one sees in large-scale data, automated data collection processes. It is possible that this small-scale research has revealed insights that can inform other similar bibliographic studies and may even inform some larger-scale studies. The two major implications for methodology stem from research issues encountered during this study: Author disambiguation and clarifying researcher domain.

Reliance on publications in top-tier journals for data in evaluating a new program. By the time an article is published in a prestigious journal, researchers may have published or presented less visible but relevant articles. Additionally, the time before any major publications result can be years beyond the grant award. In the study of cliques within this network, it was discovered that the single paper that drew the four researchers in Clique 1 was published in 2015 and the paper was funded through a 2010 NSF grant from the Division on Research on Learning on which one of these clique members was the PI and another was a Co-PI. This research was not designed to capture collaborations from any funding cycles beyond, CL, though this inadvertent discovery highlights the deep, long time connections among CL researchers.

Author disambiguation and incomplete search results._One of the first obstacles encountered in this research is name disambiguation. When two authors (or researchers) have the same name, or similar initials it is often difficult to establish which is correct name. In bibliographic and bibliometric research, this is a known problem. (Some examples of this can be found in Bordons et al.,2015; Börner et al.,2005; Xian & Madhavan, 2014.) While the issue was to be expected in citation databases, the problem of name ambiguity was also encountered in a database of NSF researchers with data coming from NSF.

Thomson Reuters' WoS core collection advertises a Distinct Author Identifier designed to help with this well-known issue. However, it quickly became apparent that the Distinct Author Identifier was held by so few researchers in this study that the feature was not useful. In the WoS Core collection, it was not always possible to know if an author with a common last name and a shared first initial was the same person as the CL researcher with the common last name and same first initial. Though every attempt was made to ensure this was the case, several searches needed to be conducted to try to catch as many citations from each author as possible. Because this study focused on co-authorships among a specific group of researchers, an incorrectly downloaded citation would be unlikely to impact the study, because another CL co-author would need to be present for that article to be included.

A related issue arises when the citations to be collected for data stem not from subject searches (which have their own set of problems) but from a known population of researchers. Beyond the issues of name ambiguity, a secondary problem of finding the complete works by a specific author arises. No one citation database or resource provides complete indexing for every citation a by every researcher. It is recommended that for the most complete coverage both major citation databases (SCOPUS and WoS) should be used, as well as Google Scholar, ResearchGate, and other online venues where researchers post their work. At the time of this writing, no single resource serves as an aggregator for the entirety of a researcher's work, so multiple resources must be consulted.

Barriers to consider before adding SCOPUS and a Google Scholar author search to the WoS citation database are cost, quality of the data, and in the case of Scholar potential downloading blockage. SCOPUS and WoS are expensive databases and while their content is not redundant, most institutions choose to purchase only one. Google Scholar appears at first glance as a wonderful source from which to obtain bibliographic data, however the data is notoriously ill-formatted. What could be a bigger barrier to using Scholar as a tool to find citations by author is the possibility of being blocked by scholar for what Google deems excessive or suspicious downloading. A researcher's CV is the most likely resource for complete publication information, however these are not conveniently located and finding these create barriers for data collection. At this moment, the ability to obtain citation information about all publications by a researcher remains a challenge.

What is needed is a unique identifier for each author or researcher. This study is not the first to encounter this problem, and others have attempted to resolve it. At this moment in time, multiple organizations are proposing unique author ID numbers. For example, Thomson Reuters introduced ResearcherID, which creates a unique author ID number that can be used in Thomson Reuters products (2015). SCOPUS, the major competitor to WoS also automatically assigns an Author ID number to anyone whose work is indexed in that database (Elsevier, 2016). A non-profit organization called ORCID (Open Researcher and Contributor ID) allows researchers to create their own profiles and provides access to some citation indexes to easily add published resources (ORCID, n.d.). The National Institutes of Health (NIH) and NSF have both begun to implement a program called SciENcv that can link to ORCID. Of these options, ORCID appears the most promising due to the ease of creation and editing of a profile, the open nature of the ORCID platform, and because ORCID is the only one not tied to a government organization or a product.

Ideally, a single entity would issue a unique author ID number that would be universally recognized, like the ISBN numbering system. In the meanwhile, the quest for a unique author ID number to help with the problem of disambiguation has resulted in the creation of at least three potential unique ID numbers. Finding out if a funding program has its own system in place to identify publications created by grantees should be a step in any co-author analysis.

Identifying researcher domains. There are some inherent difficulties in attempting to classify a researcher into a domain. At an organizational level, not every institution structures its academic departments in the same way, or has the same names, or even offers a degree with a standard name. Xien and Madavan (2014) discuss the difficulty in matching up academic departments from different institutions, due to the variety of ways in which universities structure their programs or offer unique programs. At an individual level, top level researchers tend to shift the focus of their research over time, and a simple departmental heading may not capture the complexity of a researcher's work. Furthermore, the nature of researchers is that many of them grow and shift the focus of their research across the span of their careers; this creates several issues when trying to classify a researcher into a domain. Publication lists, conference presentations and grant awards listed in Curricula Vitae provided the most current insights into the interests of a researcher at any moment in time. Table 5 shows the distribution of researcher domains among the CL grantees. Four research domains combined make up 18% of the network, were researchers allowed to place themselves into the 7 research domains used in this research, it's likely the distribution would look different.

The application of domain attributes to network data brings a wonderful richness and context to the data. Were there a way to capture researcher domains (and to account for the fluidity of researchers moving across domains), other research would surely benefit from access to this data. There are already large-scale tools in existence to which the application of domain attributes would allow easy, quick views into the domain make-up of an NSF award team with but a click of a mouse.

Recommendations for Future Research

This exploratory research provided baseline data about the structural social capital of the NSF funded CL community and insights into the diversity of research teams and their publication partner. From a broader perspective, this research is a proof of concept for using existing, publicly available data to gain insight into the structural social capital to stakeholders of interdisciplinary research communities. The low hanging fruit of any future research is to continue this project by including those researchers who have received CL related awards from NSF into the CL co-author network. NSF has coded those projects that are CL-themed with the reference code 8244. There is more pressing work to be done that could make this research method stronger and of greater value to stakeholders, though. An expansion of the co-author collaborations beyond the CL grantees would create a richer network, though it is hard to tell if this would create noise, or yield good insights. Another suggestion for stakeholders who are curious to know about how prolific a group of researchers are would be to include an attribute for solo-published articles (which could, for example be shown by node size or shading). The inclusion of more attribute data would provide a better view of diversity among an award team or a publication team. In those instances where the population is known, generated data on domain would create an accurate, current picture of the domain make-up of a community. In all emerging interdisciplinary fields of research, such as CL, further attention should be given to the terms a community uses to describe its domains and its work. Finally, engaging with the researchers to follow up on findings in this research is recommended.

Attributes. The usefulness of a co-author analysis depends almost entirely the addition of quality attribute data. Though there are limits to what a co-author analysis can reveal about the social structural capital of any research community, there is much to be gained by adding additional attribute data to this type of study. More attribute data should be collected at the award level, and at the PI/Co-PI level. For example, a request for the dates of researchers' last terminal degrees could be an indicator of how long the researcher has been active in the field. For the purposes of this study, other existing resources where the addition of domain attribute information would reveal the interdisciplinarity of research teams would be helpful within Award Search or DIA2., It would not be difficult to add award numbers (or similar) as attributes to a co-author analysis network; this addition would allow stakeholders to see what has been written by award teams.

Other attribute data specific to a funding agency or center could also be included. It would be possible, for example to add conference or webcast attendance as attribute data and see if any measure of impact can be traced to those events. Were there a way to track connections made at conferences or other professional gatherings hosted by stakeholders, this type of attribute data might reveal whether these professional networking opportunities lead to further collaborations (e.g., publications or research). Naturally, privacy protection must be paramount in any collection or requests for these kinds of attribute data. There seems to be potential for stakeholders to use social network analysis methods for formative or summative assessments when rich attribute data is available.

At the award level, it may also be appropriate to categorize multi- or interdisciplinary groups of researchers by the type of project funded. If a funding agency, research center, or interdisciplinary consortium has already created a taxonomy of the types of research done by its research community, these terms could easily be included as attributes associated with individual researchers. Applied to a network graph, research taxonomy information could reveal intersections of and bridges across different research areas of the field; this information could provide stakeholders with the justification they need to create seminars or to bring specific groups of researchers together.

Federally funded programs have an obligation to ensure diversity among grantees. It is possible that publicly available data, such as that used in this study, could contribute to diversity assessments. By collecting graduation dates, degrees, and universities attended, it would be possible to gain understanding of the academic diversity and diversity among newcomers and old-timers in a research community. (E.g., one could see the group of researchers graduated from the Northwestern Learning Science program between a certain set of dates.) Another way to look at diversity is the target population of an award. For example, CL strives to serve K to Gray (Borgman et al., 2008); a brief look at the awards seems to indicate that awards are sparser at either end of that spectrum, with a concentration of awards focused on middle school, high school, and college populations. Attribute data could be applied to look at awards targeting ability and access (e.g., several CL awards are concerned with applications for hearing impaired.) or marginalized populations would provide another lens on diversity. Diversity data could be used by stakeholders as part of an evaluation process. Additionally, though publication is a narrow way of measuring impact, adding attribute data regarding funding amount, award type[9] (if a program offers multiple awards), and duration of the award would allow stakeholders to see what types of awards generate more publications.

Domain expertise. More research is needed draw conclusions about the distribution of domain expertise among CL awards. First, high-level researchers, such as those who win federal research grants may drift away from their original doctoral training as they follow new research trends and interests. Over a five-year period, a researcher may have embraced a new research identity or claim a domain that she was only exploring earlier. To accommodate for these shifts it is recommended that stakeholders interested in identifying the research domains of their multi- and interdisciplinary communities send out short (45 second!) surveys to their researchers to get this information first hand. This sort of survey would give a controlled vocabulary of domains to be used as attributes, and in and of itself would provide insight to stakeholders about their community of researchers.

Regarding researcher domain, one final area of further work this research recommends is consideration of how to account for interdisciplinarity that lies within a single researcher. During this investigation, it became apparent that many of the researchers in the CL community had multidisciplinary backgrounds (e.g., PhDs in law and computer science or the researcher

[9] See Appendix G for information about CL Award types

with PhDs in Education and Anthropology, whose work centers around kids and media). Research by Yoon and Hmelo-Silver (2017) seem to verify this observation; they surveyed 253 members of the International Society of the Learning Sciences (ISLS) and found tremendous diversity in both the domain interests and research foci of the academics, doctoral students, academic professional staff and non-academic professional staff. In many cases, the distribution of an award to a single person could be categorized as interdisciplinary. This is not likely work that can be done on behalf of the research community writ large, but is a consideration for groups of multi- and interdisciplinary stakeholders. Future work designing surveys for rapid response and return by PI/Co-PI team members should consider that a single researcher may identify in more than one domain, or may be an expert in one domain while emerging into another.

Clarification of terms and expansion of the research. This final set of recommendations is specifically relevant to the CL community, however similar challenges likely face any new interdisciplinary or multidisciplinary field. At this moment in time it seems that many terms used within the CL community are fuzzy or contested, two of them discussed here are Learning Sciences and Cyberlearning. It is not clear what makes a researcher a learning scientist. One can be classified as being in the Learning Sciences domain by graduating from a university specifically offering that degree, but there are many integral members of the CL community who identify with and would be classified in the domain of Learning Science, but who arrived there by another path; some were researchers before the field or Learning Sciences began, and some because there are few places that provide a degree in the Learning Sciences. This ambiguity, in addition to the large number of interdisciplinary researchers mentioned above, makes domain classification within the CL community challenging. Results from a survey of 253 ISLS members illuminate the variety of domain and research expertise among those surveyed and the authors illustrate that rather than a single identity among learning scientists, many sub-communities exist (Yoon & Hmelo-Silver, 2017).

Regarding the term Cyberlearning itself, London's (2012) work highlighted the difficulties in searching for CL related grants when there is not a clear definition of what CL is, and what would constitute a CL award. One of the problems researchers encounter when searching NSF resources for cyberlearning is a lack of controlled vocabulary for cyberlearning that would make it easy to identify which pre-C:FLT NSF awards fall under the umbrella of cyberlearning (London, 2012). It's important to realize that this is not simply an NSF issue, it translates to database searches, too. (The lack of clarity around CL and CL-related terms is one reason that this project searched PI/Co-PI names in WoS, rather than keywords.). At a 2016 conference, the current Program Officer for NSF's CL program alluded to the somewhat ephemeral nature of the term Cyberlearning in his closing speech at CL 2016, noting that in 20 years, no one is going to use the term Cyberlearning, rather, it will simple be called learning (C. Hoadley, Personal communication, June. 6, 2016). Without clarification of terms used by and within the CL community, bibliographic (and any other text-based, archival) research will be hampered.

Engaging with researchers. Epistemologically, this research is grounded in a post-positivist view research can only make imperfect attempts to attain a truth (Phillips & Burbules, 2000). It has been clear throughout this process that the co-authorship lens is a narrow one; there is much more to the collaborations and communications of a group of researchers than can be revealed through applying social network analysis tools to co-author data. It is also clear that the use of secondary data, while rich with attributes, would benefit greatly from actual conversations with or survey data from the researchers themselves. Stakeholders are able to strategically engage with their researcher communities to supplement this kind of research.

Conclusion

Using social network analysis tools to make co-author networks visible holds potential for stakeholders in interdisciplinary research. It provides evidence of successful working relationships and networks that have developed through funding and can show area of dense connections and areas that might need to grow. It highlights multi-disciplinary writing teams, and can be used to set a baseline for publishing growth within a new area of research. Data for this type of research can be obtained from publicly available sources; researcher-provided data would allow for richer attribute data and more complete bibliographic data.

REFERENCES

Association of American Universities, Interdisciplinary Task Force. (2005). *Report of the Interdisciplinary Task Force.* Retrieved from http://www.aau.edu/workarea/downloadasset.aspx?id=462

Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development*, *53*(1), 86-107. http://dx.doi.org/10.1007/bf02504859

Barabâsi, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and Its Applications, 311*(3), 590-614. http://dx.doi.org/10.1016/s0378-4371(02)00736-7

Ben-David, J., & Collins, R. (1966). Social factors in the origins of a new science: The case of psychology. *American Sociological Review*, 451-465. http://dx.doi.org/10.2307/2090769

Börner, K., & Polley, D. E. (2014). *Visual insights: A practical guide to making sense of data.* Cambridge, Mass: The MIT Press

Börner, K., Dall'Asta, L., Ke, W., & Vespignani, A. (2005). Studying the emerging global brain: Analyzing and visualizing the impact of co-authorship teams. *Complexity*, *10*(4), 57-67. http://dx.doi.org/10.1002/cplx.20078

Bordons, M., Aparicio, J., González-Albo, B., & Díaz-Faes, A. A. (2015). The relationship between the research performance of scientists and their position in co-authorship networks in three fields. *Journal of Informetrics*, *9*(1), 135-144. http://dx.doi.org/10.1016/j.joi.2014.12.001

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks. 27*(1), 55-71. *http://dx.doi.org/10.1016/j.socnet.2004.11.008*

Borgatti, S. P., & Cross, R. (2003). A relational view of information seeking and learning in social networks. *Management Science*, *49*(4), 432-445. http://dx.doi.org/10.1287/mnsc.49.4.432.14428

Borgatti, S. P., Everett, M. G. and Freeman, L. C. (2002). Ucinet 6 for Windows: Software for social network analysis. [Computer software] Harvard, MA: Analytic Technologies.

Borgatti, S. P., Everett, M. G., & Johnson, J. C. (2013). *Analyzing social networks*. Los Angeles, CA: SAGE Publications Inc.

Borgman, C. L., Abelson, H., Dirks, L., Johnson, R., Koedinger, K. R., Linn, M. C., … Szalay, A. (2008). Fostering learning in the networked world: The cyberlearning opportunity and challenge a 21st century agenda for the national science foundation. *National Science Foundation, Task Force on Cyberlearning*. Retrieved from: https://www.nsf.gov/pubs/2008/nsf08204/nsf08204.pdf

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education*, (pp. 241–58). Westport, CT: Greenwood.

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, *64*(3), 351-374. http://dx.doi.org/10.1007/s11192-005-0255-6

Bransford, J., & National Research Council (U.S.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, D.C: National Academy Press. doi:10.17226/9457

Burt, R. S. (1992). *Structural holes: The social structure of competition*. Cambridge, Mass: Harvard University Press.

Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior*, *22*, 345-423. http://dx.doi.org/10.1016/s0191-3085(00)22009-1

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, *110*(2), 349-399. http://dx.doi.org/10.1086/421787

Burt, R. S. (2005). *Brokerage and closure: An introduction to social capital*. Oxford, England: Oxford University Press.

Bush, V. (1945). Science: The endless frontier. *Transactions of the Kansas Academy of Science* (1903), 231-264.

CIRCL | Center for Innovative Research in Cyberlearning. (n.d.-a). *Big ideas: Cyberlearning timeline*. Retrieved from: http://circlcenter.org/bigideas/

CIRCL | Center for Innovative Research in Cyberlearning. (n.d.-b). *Press resources*. Retrieved from: http://circlcenter.org/about/press/

CIRCL | Center for Innovative Research in Cyberlearning. (n.d.-c). I. Retrieved from: http://circlcenter.org/projects/tag-map/

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, S95-S120. http://dx.doi.org/10.1086/228943

Coleman, J. C. (1994). *Foundations of social theory*, Cambridge, Mass.: Harvard University Press.

Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. Chicago: University of Chicago Press.

Crain, R., Cooper, C., & Dickinson, J. L. (2014). Citizen science: a tool for integrating studies of human and natural systems. *Annual Review of Environment and Resources*, *39*(1), 641. http://dx.doi.org/10.1146/annurev-environ-030713-154609

Cross, R., Borgatti, S. P., & Parker, A. (2002). Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Reviews, 44*(2), 25-46. http://dx.doi.org/10.2307/41166121

DIA2. (n.d.). Retrieved from http://www.dia2.org

Dietz, J. S., Chompalov, I., Bozeman, B., Lane, E. O. N., & Park, J. (2000). Using the curriculum vita to study the career paths of scientists and engineers: An exploratory assessment. *Scientometrics*, *49*(3), 419-442. http://dx.doi.org/10.1023/a:1010537606969

De Solla Price, D. J. (1965). Networks of scientific papers. *Science 149* (3683): 510–515 doi:10.1126/science.149.3683.510

De Solla Price, D. J., & Beaver, D. (1966). Collaboration in an invisible college. American Psychologist, *21*(11), 1011-1018. http://dx.doi.org/10.1037/h0024051

Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of co-authorship and citation networks. *Journal of Informetrics*, *5*(1), 187-203. http://dx.doi.org/10.1016/j.joi.2010.10.008

Elsevier. (2016) *Manage my author profile.* Retrieved from: https://www.elsevier.com/solutions/scopus/support/authorprofile

Fields, C. (2015). How small is the center of science? Short cross-disciplinary cycles in co-authorship graphs. *Scientometrics*, *102*(2), 1287-1306. http://dx.doi.org/10.1007/s11192-014-1468-3

Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity, and creative success. *Administrative Science Quarterly*, *52*(3), 443-475.

Games, A & Squire, K. D. (2011) Searching for the fun in learning: A historical perspective on the evolution of educational video games. In S. Tobias & J. D. Fletcher (Eds.). *Computer games and instruction*. (pp 17-46) Albany, NY: State University of New York.

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

Glänzel, W. & Schubert, A. (2005). Analyzing scientific networks through co-authorship. In H. F. Moed, W. Glänzel, & U, Schmoch (Eds.), *Handbook of quantitative science and* *technology research: The use of publication and patent statistics in studies of S & T systems*. (pp. 257-276). Dordrecht: Kluwer Academic Publishers.

Godin, B. (2006). On the origins of bibliometrics. *Scientometrics*, *68*(1), 109-133. http://dx.doi.org/10.1007/s11192-006-0086-0

Gould, R. V., & Fernandez, R. M. (1989). Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological Methodology*, *19*(1), 89-126. http://dx.doi.org/10.2307/270949

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 1360-1380. http://dx.doi.org/10.1016/b978-0-12-442450-0.50025-0

Guan, J., Liu, N. (2014). Measuring scientific research in emerging nano-energy field. *Journal of Nanoparticle Research, 16*(4), 1-15. doi:10.1007/s11051-014-2356-8

Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. Riverside, CA: University of California, Riverside. Retrieved from: http://faculty.ucr.edu/~hanneman/)

Harmelen, M. (2012). Analytics for understanding research. *CETIS Analytics Series*, *1*(4), 1-41. Retrieved from http://publications.cetis.ac.uk/wp-content/uploads/2012/12/Analytics-for-Understanding-Research-Vol1-No4.pdf

Harzing, A. W. K., & Van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*, *8*(1), 61-73. http://dx.doi.org/10.3354/esep00076

Hess, D. J. (1997). *Science studies: An advanced introduction*. NYU press.

Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, *52*(2), 291-314. http://dx.doi.org/10.1023/a:1017919924342

IAIED | International AIED Society. (n.d.). About. Retrieved from http://iaied.org/about/

ISLS | International Society of the Learning Sciences. (2015). *Creation of the society*. Retrieved from: https://www.isls.org/about/creation-of-the-society

Jeanes, E., Loacker, B., & Śliwa, M. (2014). Researcher collaboration: Learning from experience. In E. Jeanes, & T. Huzzard (Eds.), *Critical management research: Reflections from the field.* (pp. 41-61). London, UK: SAGE Publications Ltd. doi:10.4135/9781446288610.n3

Jonassen, D. H. (1992). What are cognitive tools? In P. A. M. Kommers, D. H. Jonassen, & J. T. Mayes (Eds.) *Cognitive tools for learning* (pp. 1-6). Berlin, Germany: Springer.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1-18. http://dx.doi.org/10.1016/s0048-7333(96)00917-1

Klavans, R., & Boyack, K. W. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology*, *60*(3), 455-476. doi:10.1002/asi.20991

Knoke, D., & Yang, S. (Eds.). (2008). *Social network analysis.* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781412985864

Krackhardt, D. (1999). The ties that torture: Simmelian tie analysis in organizations. *Research in the Sociology of Organizations*, *16*(1), 183-210.

Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation, 11*(1), 3-15. http://dx.doi.org/10.3152/147154402781776961

Ledford, H. (2015). How to solve the world's biggest problems*. Nature*, *525*(7569), 308-311. doi:10.1038/525308a. Retrieved from: http://www.nature.com/news/how-to-solve-the-world-s-biggest-problems-1.18367s.

Library of Congress. (2008). *Library of Congress policy statements: Technical reports, working papers and preprints.* Retrieved from: https://www.loc.gov/acq/devpol/techrept.pdf

Lin, N. (1999). Building a network theory of social capital. *Connections 21*(1). 28-51.

Lin, N. (2001a). The position generator: Measurement techniques for investigations of social capital. In N. Lin, K.S. Cook & R.S. Burt (Eds.), *Social capital: Theory and research*. (pp. 57-81) New York: Aldine de Gruyter.

Lin, N. (2001b). *Social capital: A theory of social structure and action*. Cambridge, Mass: Cambridge University Press.

Lingo, E. L., & O'Mahony, S. (2010). Nexus work: Brokerage on creative projects. *Administrative Science Quarterly*, *55*(1), 47-81. http://dx.doi.org/10.2139/ssrn.1410764

Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing & Management*, *41*(6), 1462-1480. http://dx.doi.org/10.1016/j.ipm.2005.03.012

London, J. (2012). *Exploring cyberlearning through a NSF lens.* Paper presented at 2012 ASEE Annual Conference and Exposition in San Antonio, Texas. Retrieved from http://www.asee.org/public/conferences/8/papers/4153/view

Long, J. C., Cunningham, F. C., & Braithwaite, J. (2013). Bridges, brokers and boundary spanners in collaborative networks: A systemic review. *Health Services Research, 13*(158), 1-13. http://dx.doi.org/10.1186/1472-6963-13-158

Leydesdorff, L. & and Milojevič, S. (2015). Scientometrics. In J. D. Wright (Editor-in-chief), *International encyclopedia of social and behavioral sciences: Vol. 21* (322-327). Oxford: Elsevier.

Macarthur Foundation (2017). *Research networks*. Retrieved from https://www.macfound.org/networks/

McFadyen, M. A., & Cannella, A. A. (2004). Social capital and knowledge creation: Diminishing returns of the number and strength of exchange relationships. *Academy of Management Journal*, *47*(5), 735-746. http://dx.doi.org/10.2307/20159615

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27415*. https://doi.org/10.1146/annurev.soc.27.1.415

Milesi, C., Brown, K. L., Hawkley, L., Dropkin, E., & Schneider, B. L. (2014). Charting the impact of federal spending for education research: A bibliometric bpproach. *Educational Researcher*, *43*(7), 361-370. http://dx.doi.org/10.3102/0013189x14554002

Milgram, S. (1967). The small world problem. *Psychology Today 1*(1), 60-67. http://dx.doi.org/10.1037/e400002009-005

Molenda, M. (2007). Historical foundations. In M.J. Spector (Ed.), *Handbook of research on educational communications and technology* (pp. 3-20). London, UK: Routlage.

Mooghali, A., Alijani, R., Karami, N., & Khasseh, A. A. (2012). Scientometric analysis of the scientometric literature. *International Journal of Information Science and Management (IJISM)*, *9*(1), 19-31. Retrieved from http://ijism.ricest.ac.ir/index.php/ijism/article/download/89/80

Nadworny, E. (2014, October 22). The Slide Rule: A Computing Device That Put A Man On The Moon. npr Ed. Retrieved from:

http://www.npr.org/sections/ed/2014/10/22/356937347/the-slide-rule-a-computing-device-that-put-a-man-on-the-moon

National Academy of Sciences, National Academy of Engineering, & Institute of Medicine. (2005). *Facilitating interdisciplinary research*. Washington, DC: The National Academies Press. https://doi.org/10.17226/11153.

National Science Foundation. (2011). *FY 2012 budget request to Congress.* Arlington, Virginia. Retrieved from: https://nsf.gov/about/budget/fy2012/pdf/fy2012_rollup.pdf

National Science Foundation. (2014a). *Science and engineering indicators 2014*. Arlington, Virginia. Retrieved from: http://www.nsf.gov/statistics/seind14/index.cfm/chapter-4/c4s1.htm#s2

National Science Foundation. (2014b). *Strategic plan for 2014-2018*. Arlington, Virginia. Retrieved from: https://www.nsf.gov/pubs/2014/nsf14043/nsf14043.pdf

Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404-409. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC14598/

Newman, M. E. (2004). Co-authorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*(suppl 1), 5200-5205. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC387296/

Obstfeld, D. (2005). Social networks, the tertius iungens orientation, and involvement in innovation. *Administrative Science Quarterly*, *50*(1), 100-130.

ORCID. (n.d.). *About*. Retrieved from: http://orcid.org/about

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Lanham, MD: Rowan & Littlefield Publishers, Inc.

Pinch, T. J., & Bijker, W. E. (1987). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. *Social Studies of Science*, *14*(3), 339-441. https://doi.org/10.1177/030631284014003004

playableUCSC. (November 11, 2013). *Janet Kolodner -- Cyberlearning: Transforming Education (Media Systems #13)*. Retrieved from https://youtu.be/Ln4Iuvhip1s?list=PLdLVzw_imoZhDA6arc1YQxUUY1B22WDL5

Portes, A. (1998). Social capital: Its origins and applications in modern society. *Annual Review of Sociology*, 24, 1-24. https://doi.org/10.1146/annurev.soc.24.1.1

Putnam, R. D. (2001). Bowling alone: The collapse and revival of American community. New York: Simon and Schuster. doi:10.4324/9781912282319

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., & Kafai, Y. (2009). Scratch: programming for all. *Communications of the ACM*, 52(11), 60-67. http://dx.doi.org/10.1145/1592761.1592779

Rodriguez, M. A., & Pepe, A. (2008). On the relationship between the structural and socioacademic communities of a co-authorship network. *Journal of Informetrics*, *2*(3), 195-201. doi:10.1016/j.joi.2008.04.002

Rogers, E. M. (2003). *Diffusion* of innovations (5th ed.). New York: Free Press.

Roschelle, J. M., Pea, R. D., Hoadley, C. M., Gordin, D. N., & Means, B. M. (2001). Changing how and what children learn in school with computer-based technologies. *The Future of Children*, *10*(2), 76-101. http://dx.doi.org/10.2307/1602690

Schlager, M. S., Farooq, U., Fusco, J., Schank, P., & Dwyer, N. (2009). Analyzing online teacher networks cyber networks require cyber research tools. *Journal of Teacher Education*, *60*(1), 86-100. http://dx.doi.org/10.1177/0022487108328487

Scott, J. (2013). *Social network analysis* (3rd ed.). Thousand Oaks, CA: SAGE Publications, Inc.

Simmel, G. (1950). *The sociology of Georg Simmel.* (K. J. Wolff, Trans.) Glencoe, IL: Free Press.

Singer, P. L. (2014). *Federally supported innovations: 22 examples of major technology advances that stem from federal research support.* Washington, D.C.: The Information Technology and Innovation Foundation. Retrieved from: http://www2.itif.org/2014-federally-supported-innovations.pdf

Stigler, J. W. (1984). "Mental abacus": The effect of abacus training on Chinese children's mental calculation. *Cognitive Psychology*, *16*(2), 145-176. doi:10.1016/0010-0285(84)90006-9

Strober, M. (2015). *Interdisciplinary conversations: Challenging habits of thought*. Stanford, CA: Stanford University Press.

Technology Innovation Program. (2010). *A guide for preparing and submitting white papers to the technology innovation program*. Retrieved from: https://www.nist.gov/sites/default/files/documents/tip/wp/guide_for_white_papers.pdf

Thomson Reuters. (2014) *Web of Science Core Collection*. Retrieved from http://thomsonreuters.com/content/dam/openweb/documents/pdf/scholarly-scientific-research/fact-sheet/web-of-science-core-collection.pdf

Thomson Reuters. (2015) *ResearcherID*. Retrieved from http://www.researcherid.com/Home.action?returnCode=ROUTER.Unauthorized&SrcApp=CR&Init=Yes

Thomson Reuters. (2016). *2015 Journal Citation Reports® Social Sciences Edition.* Retrieved from www.thomsonreuters.com

Van Raan, A. F. (2005). Measuring science. In *Handbook of quantitative science and technology research* (pp. 19-50). Springer Netherlands. Retrieved from http://www.researchgate.net/profile/Thed_Van_Leeuwen2/publication/227039230_Descriptive_Versus_Evaluative_Bibliometrics/links/00b7d51933e7fd186a000000.pdf#page=25

Wagner, C. S. (2008). *The new invisible college: Science for development*. Washington, D.C: Brookings Institution Press.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., & Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, *5*(1), 14-26. https://doi.org/10.1016/j.joi.2010.06.004

Xian, H., & Madhavan, K. (2014). Anatomy of scholarly collaboration in engineering education: A big-data bibliometric analysis. *Journal of Engineering Education*, *103*(3), 486-514. doi:10.1002/jee.20052

Yin, L. C., Kretschmer, H., Hanneman, R. A., & Liu, Z. Y. (2006). Connection and stratification in research collaboration: An analysis of the COLLNET network. *Information Processing & Management*, *42*(6), 1599-1613. http://dx.doi.org/10.1016/j.ipm.2006.03.021

Yoon, S. A., Hmelo-Silver, C. E. (2017). What do learning scientists do? A survey of the ISLS membership. *Journal of the Learning Sciences*, *26*(2), 167-183. http://dx.doi.org/10.1080/10508406.2017.1279546

APPENDIX A

CL Project Tag Index From circlcenter.org

Accessibility and technology (3)

Analytics/data mining (19)

Augmented reality/immersive environments for learning (13)

Citizen science (13)

Collaborative and/or participatory learning (34)

Communities for learning (18)

Computational thinking (10)

Computer and information science (25)

Cyber-enhanced/computer-assisted assessments (11)

Data science and visualization(16)

Embodied learning (6)

Engineering (15)

Formative Assessment (9)

Games and virtual worlds (23)

Identity in learning (5)

Industry (3)

Informal learning (22)

Inquiry learning (23)

Instruction delivery platforms(7)

Intelligent tutors and tools (10)

Interactive learning materials(19)

Learning in technology-rich environments (29)

Learning through data (15)

Making learning tangible (9)

Mathematics (19)

Mobile learning (10)

Modeling and simulation (19)

MOOCs (3)

Peer production (4)

Personalized learning (10)

Professional development (8)

Robotics (4)

Science (36)

Social sciences (17)

Statistical methods or research design (13)

Virtual and remote laboratories(5)

APPENDIX B

Sixteen Consensus Disciplines Identified by Klavans & Boyack (2009)

1. M - Mathematics
2. CS - Computer science
3. P - Physics
4. PC - Physical chemistry
5. C - Chemistry
6. E - Engineering
7. G - Earth sciences (geoscience)
8. BC - Biochemistry
9. B - Biology
10. I - Infectious disease
11. MD - Medical specialties
12. HS - Health services
13. N - Brain research (neuroscience)
14. PS - Psychology/psychiatry
15. SS - Social sciences
16. H - Humanities

APPENDIX C

Assigning Domain Codes

Departmental Affiliation / Degree Domain	**Domain Assignment**
Department affiliation and domain of last degree match	Use the matched domain.
Departmental affiliation is ambiguous (i.e., Researcher works at a research center or is director of an interdisciplinary center)	Use domain of last degree
Departmental affiliation and domain of last degree don't match.	If domain of last degree is Learning Sciences, assign that. Otherwise: If research areas of interest are available, select the departmental affiliation or degree domain that most closely matches the research interests. If research areas of interest are not available, assign the departmental affiliation

Domain Codes

First Coding Cycle	**Second Coding Cycle**	**Researchers**
Learning Sciences	Learning Sciences	58
Computer Science	Computer Science	22
Human Computer Interaction	HCI	6
Sustainability Design Information Science	Interdisciplinary Studies	4
Meteorology Mathematics Engineering Bioengineering	Science/Math/Engineering	6
Cognitive Science Artificial Intelligence	AI/Cog Sci	3
Education Communication Social Psychology	Humanities	11

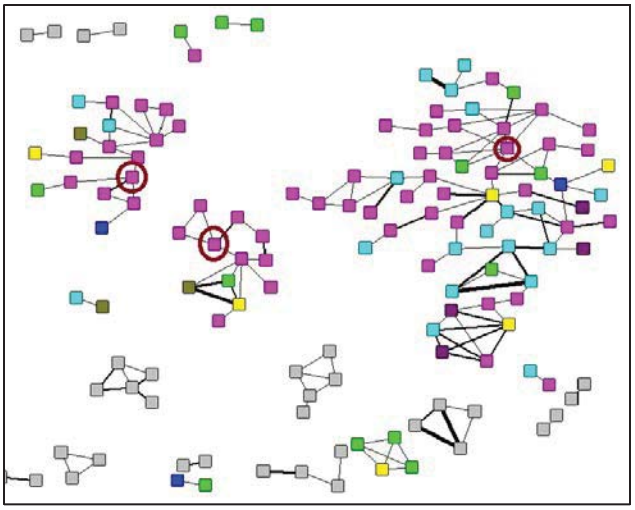
APPENDIX D

37 Nodes and Their Brokering Positions Within the Network

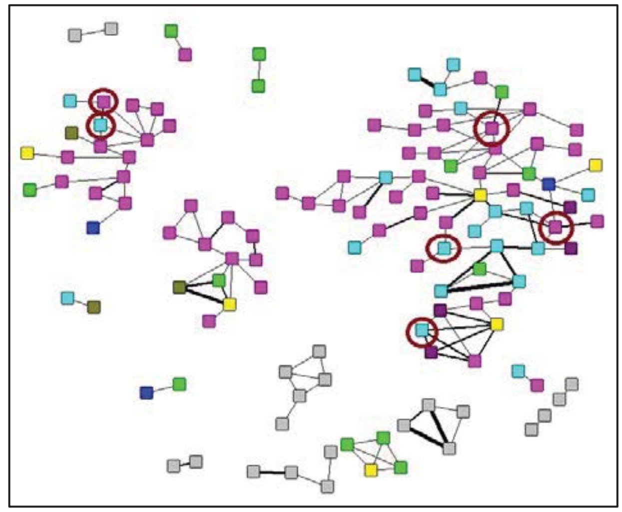
Node Number	Domain	Coordinator	Gatekeeper	Representative	Consultant	Liaison	Total
282	Cog Sci/AI	1	0	0	0	0	1
102	Cog Sci/AI	0	0	0	0	3	3
52	CS	2	0	0	0	0	2
224	CS	1	0	0	0	0	1
260	CS	1	0	0	0	0	1
14	CS	0	2	0	0	1	3
389	CS	0	0	1	1	0	2
80	CS	0	0	0	1	0	1
366	CS	0	0	0	1	0	1
103	HCI	0	0	0	0	2	2
292	LS	3	2	0	0	0	5
63	LS	3	0	2	0	0	5
190	LS	3	0	0	0	0	3
399	LS	3	0	0	0	0	3
192	LS	2	0	0	0	0	2
326	LS	1	2	0	0	0	3
60	LS	1	1	0	0	1	3
37	LS	1	0	2	0	0	3
118	LS	1	0	0	0	0	1
176	LS	1	0	0	0	0	1
207	LS	1	0	0	0	0	1
222	LS	1	0	0	0	0	1
234	LS	1	0	0	0	0	1
373	LS	1	0	0	0	0	1
54	LS	0	1	0	0	0	1
110	LS	0	1	0	0	0	1
297	LS	0	0	2	0	1	3
337	LS	0	0	1	1	1	3
187	LS	0	0	1	0	0	1
202	LS	0	0	1	0	0	1
367	LS	0	0	1	0	0	1
89	LS	0	0	0	1	4	5
105	LS	0	0	0	0	1	1
277	LS	0	0	0	0	1	1
147	LS	0	0	0	1	0	1
141	Sci/Math/Eng	1	0	0	0	0	1
189	Sci/Math/Eng	1	0	0	0	0	1

APPENDIX E

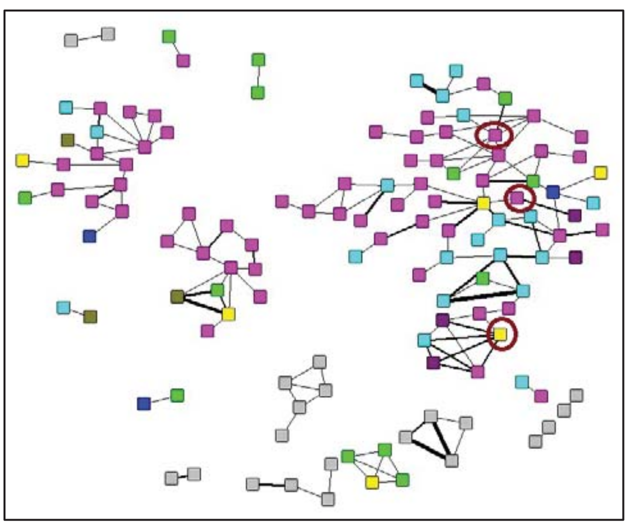
Examples of Brokering Roles Within the Network



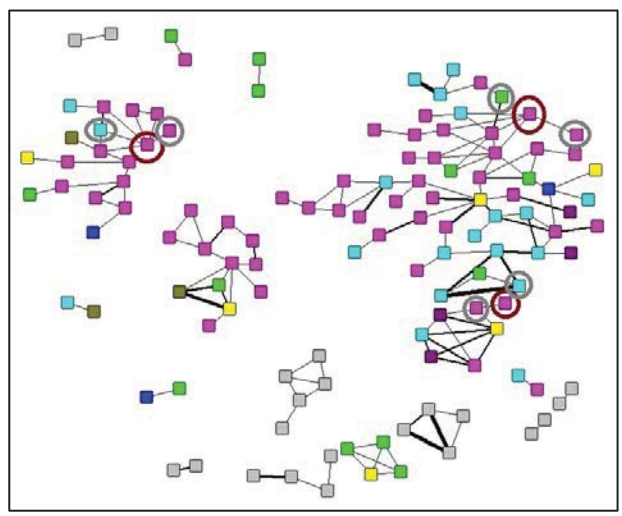
Three examples of nodes that occupy a Coordinating position within the network.
(A->A->A)



All six instances of nodes that occupy the Consultant role within the network.
(B->A->B)



Three examples of nodes occupying the Liason role within the network.
(B->A->C)



Three examples of nodes occupying Gatekeeper/Representative Roles.
(A->A->B / B->A->A)

Nodes in Gatekeeper/Representative Roles are circled in red, grey circles indicate the other nodes involved in these groupings.

APPENDIX F

Institutional Review Board Letter of Approval: Pepperdine University

PEPPERDINE

Pepperdine University
24255 Pacific Coast Highway
Malibu, CA 90263
TEL: 310-506-4000

NOTICE OF APPROVAL FOR HUMAN RESEARCH

Date: January 20, 2016

Protocol Investigator Name: Christine Forte

Protocol #: 15-10-070

Project Title: SOCIAL CAPITAL AND EXPERTISE IN A NEWLY-FORMED RESEARCH COMMUNITY: A CO-AUTHOR ANALYSIS

School: Graduate School of Education and Psychology

Dear Christine Forte:

Thank you for submitting your application for exempt review to Pepperdine University's Institutional Review Board (IRB). We appreciate the work you have done on your proposal. The IRB has reviewed your submitted IRB application and all ancillary materials. Upon review, the IRB has determined that the above entitled project meets the requirements for exemption under the federal regulations 45 CFR 46.101 that govern the protections of human subjects.

Your research must be conducted according to the proposal that was submitted to the IRB. If changes to the approved protocol occur, a revised protocol must be reviewed and approved by the IRB before implementation. For any proposed changes in your research protocol, please submit an amendment to the IRB. Since your study falls under exemption, there is no requirement for continuing IRB review of your project. Please be aware that changes to your protocol may prevent the research from qualifying for exemption from 45 CFR 46.101 and require submission of a new IRB application or other materials to the IRB.

A goal of the IRB is to prevent negative occurrences during any research study. However, despite the best intent, unforeseen circumstances or events may arise during the research. If an unexpected situation or adverse event happens during your investigation, please notify the IRB as soon as possible. We will ask for a complete written explanation of the event and your written response. Other actions also may be required depending on the nature of the event. Details regarding the timeframe in which adverse events must be reported to the IRB and documenting the adverse event can be found in the *Pepperdine University Protection of Human Participants in Research: Policies and Procedures Manual* at community.pepperdine.edu/irb.

Please refer to the protocol number denoted above in all communication or correspondence related to your application and this approval. Should you have additional questions or require clarification of the contents of this letter, please contact the IRB Office. On behalf of the IRB, I wish you success in this scholarly pursuit.

Sincerely,

Judy Ho, Ph.D., IRB Chairperson

cc: Dr. Lee Kats, Vice Provost for Research and Strategic Initiatives

APPENDIX G

Cyberlearning

As mentioned in the previous chapter, Cyberlearning is a dynamic area of research and the definition of Cyberlearning is evolving. In an early press conference about the CL program, the thenProgram Officer Janet Kolodner explained that NSF views Cyberlearning is "using any technology to amplify, expand, and provide learning opportunities that would otherwise not be available" (playableUCSC, 2013, 1:05 min). In order to understand how Cyberlearning came to be, it is important to consider the evolution of tools and technologies to enhance learning over time. Technologies that extend human abilities have been around since the beginning of humankind. It is the creation and use of tools to extend physical and mental capabilities that distinguishes higher order thinking; cognitive tools are those which "compliment and extend the mind" (Jonassen, 1992, p. 3). Roy Pea (1985) uses the term cognitive technology to describe "any medium that helps transcend the limitations of the mind, such as memory, in activities of thinking, learning, and problem solving" (p. 168); Pea names a diverse range of cognitive technologies that range from written language to computers. The abacus, which is still used in some classrooms and shops around the world, is an early example of a cognitive technology used to mediate learning (Jonassen, 1992; Pea, 1985; Stigler, 1984). Throughout the history of formal education, technologies of the time found their way into classroom and changed the way people taught and learned. Pre-industrial revolution classroom teachers introduced slates and pencils became tools in the classroom, followed by slide projections, audio, and video (Molenda, 2007). The first slide rule was built in England in the 1600's; this technology was heavily used in mathematics, science, and engineering up until 1972, when the introduction of Hewlett Packard's hand-held calculator rapidly took over as the tool of choice (Nadworny, 2014).

As computing technologies became more powerful and more commonly used in research, learning scientists began to consider ways that these tools could effectively be brought into classroom learning (Bransford, 2000; Roschelle, Pea, Hoadley, Gordin, & Means, 2001). Some examples of early Cyberlearning in the classroom are the choice-based, historical game Oregon Trail and the programming game Logo that gained popularity in primary school classrooms in the mid-1970s. Oregon Trail began as a text-based adventure type game designed to teach students about the American migration to the West, while Logo used natural language and a cartoon turtle to teach early programming skills (Games & Squire, 2011). Two recent popular examples of Cyberlearning in the classroom are Quest Atlantis, a 3-D multiplayer game that immerses players in a quest to solve global problems (Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005) and Scratch, a virtual community that allows children to upload and share their projects they created using Scratch's visual programming tools (Resnick et al., (2009). The newest technologies of any particular point in time tend to trickle into the classroom; this is particularly true when the technologies can be shown to offer affordances in learning, decrease cognitive load, or allow student to experience simulations or manipulate data in ways that offer genuine experiences (Jonassen, 1992). However, Cyberlearning is not just the technology tools, but the tools combined in new ways grounded in understanding of how people learn (Bransford, 2000).

The National Science Foundation (NSF) officially adopted the term Cyberlearning in - 2008 in the task force report "Learning in the Networked World" (Borgman et al). In 2015, Cyberlearning is still a young concept, the definition remains fluid and it is difficult to predict whether Cyberlearning will one day be classified as a discipline, if it will be an approach to learning, or how it will be viewed in the next five to ten years. The newness and dynamism of Cyberlearning is part of the appeal of this research project; it is an opportunity to capture a snapshot in the growth of a phenomenon. In order to understand who the population of Cyberlearning researchers are, and why NSF started a new program about Cyberlearning, it is important to understand some of the history of NSF and its role in technology in learning.

History of technology and cyberlearning at NSF. In 1944 President Franklin D. Roosevelt requested that Vannevar Bush, Director of the United States Office of Scientific Research and Development (OSRD), propose a way to ensure a peacetime continuation of the research and development that flourished under OSRD during World War II (Bush, 1945.). Bush's proposal led to the establishment of the National Science Foundation in 1950 with the mission "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense; and for other purposes" (National Science Foundation, 2014b, p. 3). and the vision "A Nation that creates and exploits new concepts in science and engineering and provides global leadership in research and education" (National Science Foundation, 2014b, p. 3). Cyberlearning, in many ways, is a natural outcome of the commitment to scientific discovery and education that has been part of NSF since its inception.

The cutting edge of technology found in a country at any particular point in time depends in large part on that government's prioritization and funding of research. In the United States, technology is pushed forward by a mix of public and private research and development investments that are carried out by Universities, the private sector, Federal Agencies and

Federally Funded Research and Development Centers (National Science Foundation, 2014a). Many current technologies that are a seamless part of life in the United States in in 2015 stemmed from federal funding. For example, while studying at Stanford, Sergey Brin and Larry Page conducted research funded with federal dollars that resulted in the PageRank search algorithm that ultimately launched the search engine Google (Singer, 2014). Other examples include closed captioning, Global Positioning Software (GPS), cell phone technologies, Magnetic Resonance Imaging (MRI), and even the backbone of the Internet (Singer, 2014). The commitment of NSF to teaching and learning in science and engineering help explain how NSF funded technology innovations became integrated with education and learning.

In much the same way that words live in common usage before dictionaries codify them, NSF funded projects that included Cyberlearning research predate the introduction of the term Cyberlearning by the NSF Task Force on Cyberlearning. Jeremi London's (2012) research into the NSF Department of Undergraduate Education's (DUE) projects from 2002-2012 found 866 awards totaling $100.6 M that met her search criteria for the term Cyberlearning. While 2002 was early in the development of Cyberlearning research, London noticed an increase in the projects that met her criteria starting in 2005-2006, well before NSF published reports on Cyberlearning or CyberInfrastructure. When London analyzed these early, pre-CL grants, she identified six forms of Cyberlearning: Remote access to an authentic or virtual environment; Online communities exclusively for educational purposes; Learning Management Systems, Distance education courses; Repositories of interactive resources; Games. In 2015, post-CL, the Cyberlearning research community has identified 36 different tags to describe CL awards (CIRCL, n.d.-c). While it is possible to see an evolution in the number of Cyberlearning programs and in the creation of a controlled vocabulary, Cyberlearning remains in a nascent stage of growth.

The definition of what Cyberlearning is (and is not) also continues to evolve. The NSF Cyberlearning Task Force defined Cyberlearning as “learning that is mediated by networked computing and communications technologies” (Borgman et al., 2008, p. 10). London (2012p. 3) cited several definitions from other researchers and used the following as the working definition for her research: “Teaching and learning that is mediated by use of technology and networks”. It is not yet clear if Cyberlearning will become a “field of study” (NSF, 2011, p. 218) or “a movement to tackle important problems of learning by investigating synergies between technology of the future and scientific understanding of how people learn” (CIRCL, n.d.-b, Description section, para 1) For the purposes of this study, it is not necessary to determine an exact definition of Cyberlearning or to decide if Cyberlearning can actually be considered a field or an approach teaching and learning, an amalgam of the two, or something else entirely.

Independent of NSF, the formation of scholarly organizations and journals helped bring together like-minded scholars to talk about teaching and learning with technology; in considering how CL came to be a program at NSF, these developments must also be considered. The International Artificial Intelligence in Education Society (IAIED) began in 1997 to promote “rigorous research and development of interactive and adaptive learning environments for learners of all ages, across all domains” (IAIED, n.d., para 1). IAIED publishes quarterly issues of the interdisciplinary International Journal of Artificial Intelligence in Education and hosts a conference every two years (since 1989). Janet Kolodner founded the peer-reviewed, multidisciplinary Journal of the Learning Sciences (JLS) in 1994 to showcase research on learning and education (CIRCL, n.d.-b). Journal of the Learning Sciences has grown to be influential; Thomson Reuters' 2015 Journal Citation Reports® ranked JLS 7 out of 57 Psychology, and 5 out of 231 in Education & Educational Research (Thomson Reuters,2016). In 2002, following the creation of JLS, the International Society for the Learning Sciences (ISLS) was founded by Chris Hoadley, Janet Kolodner, and Tim Koschmann with the goal to "advance the sciences and practices of learning across informal and formal learning spaces" (ISLS, 2015). NSF selected Janet Kolodner as the Program Officer for the Cyberlearning: Transforming Education solicitation and selected Chris Hoadley as the Program Officer for the second Cyberlearning program: Cyberlearning and the Future Learning Technologies.

CL solicitations. In 2009, an advisory committee to the NSF's Directorate for Education and Human Resources (EHR) along with the Office of Cyberinfrastructure recommended that NSF establish a program to create Cyberlearning as a field of study (NSF, 2011). NSF opened the solicitation for Cyberlearning: Transforming Education (C:TE) in 2011; funding for this program spanned fiscal years 2011 and 2012, with $30M allocated for this time period.

NSF's (2010) original 2011 C:TE solicitation outlines three goals:

1. To better understand how people learn with technology and how technology can be used productively to help people learn, through individual use and/or through collaborations mediated by technology;
2. To better use technology for collecting, analyzing, sharing, and managing data to shed light on learning, promoting learning, and designing learning environments; and

3. To design new technologies for these purposes, and advance understanding of how to use those technologies and integrate them into learning environments so that their potential is fulfilled.

From the first C:TE solicitation (NSF, 2010):

- Projects funded through Cyberlearning will typically be interdisciplinary, with the research team including members with the full range of expertise needed for success, in areas Ösuch as human learning, engagement, technology design, technology integration, education, and human-centered computing, as well as expertise in the nature of the targeted learning environment, the technologies being investigated, and carrying out data collection and analysis (Project Teams, para 1).
- The second C:TE solicitation clarified the need for interdisciplinary teams (NSF, 2011):
- It is expected that all EXP, DIP, INDP, and CAP projects funded through Cyberlearning will have interdisciplinary expertise. The project team (including PIs, senior personnel and supporting investigators, postdocs, advisory-board members, and others) should be appropriate for addressing proposed technological and research goals. Each team is expected to carry out the data collection and analysis necessary to evaluate and refine their innovation and answer their research questions. Teams should be formed accordingly (Section II, Interdisciplinary teams.)
- Furthermore, the second C:TE solicitation required that proposal include a collaboration plan that would demonstrate how the combined efforts of the team would be more effective than the sum of their parts (NSF, 2011).

The following information was taken from the initial C:TE solicitation (NSF, 2010): To meet these goals, the initial C:TE program solicitation offered three levels of funding: Exploration Projects (EXP); Design and Implementation Projects (DIP), and Integration and Deployment Projects (INDP). C:TE awards assume interdisciplinary teams will be needed in order to study a specific technology and its impact on learning in a specific environment. EXP awards receive the lowest level of funding and the shortest time to complete their work: up to $550k over 2 or 3 years. The EXP format encouraged the implementation of "novel or innovative" technologies and required the formative analyses of these projects to focus on usability and sustainability of a technology in the service of C:TE goals. EXP teams could include only a single Primary Investigator. In the second tier of awards, DIPs could be awarded up to $1.35M, which could have 4 or 5 years of time for the work. DIP winners were required to introduce a specific innovation into a specific learning environment (formal or informal). DIP researchers were required to plan formative assessments and report what was learned about the use and usability of the innovation in the learning environment. DIP project teams are larger than EXP and must include stakeholders or other who hold leadership positions in the environment to be studied. The third tier, INDP awards, was to take established cyberlearning technologies and introduce them into formal or informal learning environments in order learn about ways to scale the use of the technology or how to better integrate a promising technology. One of the major differences between DIP and INDP awards is that INDP focus on usability, feasibility, and sustainability at scale. In order to win INDP awards, researchers needed to demonstrate that the proposed technology or technologies in their grant application showed promise of being able to scale. These awards were designed to incorporate partnerships with schools or other organizations that would allow implementation of a technology into an entire system and were required to be deployed in a variety of settings. To accommodate the larger scope of the awards, IDNP awards were granted up to $5M over up to 5 years. The C:TE community of researchers therefore consists of a mix of researchers who are exploring the feasibility of new tools and technologies to support learning and those who are are working to scale promising technologies. Project teams are of varying size; DIP and INDP awards will likely consist of a broader group of researchers than are present on the awards. It is possible that non-PI/Co-PI project collaborators could collaborate with PI/Co-PIs by co-authoring or co-presenting about the project. Because INDP and DIP projects span over multiple years, it is possible that the publication cycle of articles stemming from these projects will be slower than that of EXP projects.