EXPLORING HOW TECHNOLOGY MEDIATES THE TYPES OF RELATIONSHIPS FORMED IN SOCIOTECHNICAL SYSTEMS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

COMMUNICATION AND INFORMATION SCIENCES

MAY 2012

By

Kar-Hai Chu

Dissertation Committee:

Daniel D. Suthers, Chairperson

Devan Rosen

Scott Robertson

Rich Gazan

Jenifer Winter

Sun-Ki Chai

Acknowledgements

I would like to thank several persons and organizations for their assistance and support.

The National Science Foundation Virtual Organizations as Sociotechnical Systems program, grant #0943147, 2010-2013, PI Dan Suthers.

The members of my committee, who I thought would be very like-minded at the start, but ended up with five very different perspectives of my work.

The members of LILT, especially Nathan Dwyer, for all of his support with software, and Viil Lid, for help with several of the figures used.

SRI International, in particular, Patti Schank for access and interpretations of the Tapped In data (and a most accommodating host), and Judi Fusco for our many discussions.

And to my advisor and friend, Dan Suthers, for our many discussions, for teaching me how to become a researcher, for believing in what I was doing, and for the… persistent… editing of many of my writings. I will always consider you ‘ohana.

Abstract

This work presents an exploratory study of how technology mediates the different types of relationships that are formed in sociotechnical systems. More people each day are connecting with each other through social networks, online communities, and other forms of virtual environments. Whether for education, information seeking, friendship, professional work, or other reasons, diverse technology mediated relationships are being formed. This study explores the idea that these relationships are influenced by the affordances that technology provides. When a person navigates through a sociotechnical system, how they interact with other users can depend upon the mediating artifacts provided by the system. The resulting relationships that are built on these interactions are therefore reflected by the technology. This work offers a framework for understanding how technology, user interactions, and user relationships are connected within a sociotechnical system, and uses this framework to uncover the kinds of interactions that take place in such systems, the relationships that are constituted by these interactions, and the influence of technology on these processes. Implications are drawn for how system designs can be improved to increase sociotechnical capital.

Table of Contents

List of Tables

List of Figures

1. Introduction

Information and communication technologies (ICT's) have become part of our everyday lives. As ICT's continue to evolve, the number of people on the Internet continues to grow larger. The emergence of Web 2.0 has further fueled the progress, affording the casual user a more active and participatory role in all aspects of the online world (Beer & Burrows, 2007). This is especially true for social network sites. Sites that began as small communities (e.g. MySpace for independent music, Facebook for Ivy League students) have grown broadly; niche communities have also developed rapidly, specializing in categories such as business (e.g. LinkedIn), education (e.g. Math Forum), and multimedia videos (e.g. YouTube) (Boyd & Ellison, 2007). There is a strong potential for these sociotechnical systems to create value and social capital for its users (Resnick, 2002), but there are still many questions regarding how the usage affects users, the different types of interactions and relationships that are formed within their virtual boundaries, and how such interactions are affected by the design of the environment. It is clear that the relationship between the social and the technical realms is important (Kling, 2007), and so having a better understanding of how the two are intertwined can help inform us on the best methods to build successful sociotechnical systems.

Online technology has changed dramatically since the early 1990's. Where "Web 1.0" offered mostly static web pages that must be viewed by a PC-based browser and contained content that was created by developers and programmers, Web 2.0 delivers dynamic, interactive content that can be created by anyone with basic computer skills and be viewed from almost any device (Beer & Burrows, 2007). Blogs, wikis, and social network sites are some of Web 2.0's many applications. User content contributions have become increasingly easier to create; setting up a blog requires only a few minutes of time and filling out a simple form on popular blog sites such as LiveJournal or WordPress. As the variety and quantity of content increases, more people continue to be drawn online. The large number of people going online for friendship and support on social network sites or information on Q&A/wiki sites testifies to the Internet's ease of access and potential to be a great source of information and resource exchange (Ridings & Gefen, 2004).

The general goal of this research is to *study and understand how technology mediates the different types of relationships that are formed in sociotechnical systems*. Social network sites have demonstrated that people can and do form and sustain successful relationships online. Even pre-Web 2.0 studies had begun investigating how the Internet could create active communities (Galston, 2000); sites such as Facebook are a realization of some of the early speculations. But research into how the technology affordances offered by Web 2.0 systems mediate online relationships has been sparse. More specifically, we need a better understanding of how technology mediates the kinds of interactions that occur online. The research needs to support computer scientists and social scientists alike to analyze large, complex sociotechnical systems and find out what individuals and groups are doing. The tight bond between the social and technical requires solutions and answers that address both areas of research (Licoppe & Smoreda, 2005). We need a framework and methods that can trace user activity in a sociotechnical system, identify different forms of interaction, and help us understand how the technology mediates the relationships that are created.

1.1. Research Motivations

The goal of this research is to study and understand the connection between relationships that people have with each other, the interactions that occur, and the digital artifacts that mediate such interactions. A major source of inspiration is the work of (Licoppe & Smoreda, 2005). Their research examined the relationship between social networks and ICT's used for communication. These components affect the interactions between actors by providing resources in support, but also by providing constraints. Licoppe and Smoreda studied empirical data of how the announcement of the birth of a child was conducted, the type of media used for communication, and the relationship the new mother had with each person she contacted. They also used data showing the changes in how people communicated after someone had just moved (geographically). In studying different modes of communication, and other factors such as changes in frequency, they focused on two modes that affect relationships: intermittent presence and connected presence. The results showed that people will constantly re-negotiate whom they communicate with, and how they choose to do so. As such, the relationship between social networks and technology is not one-sided, but rather cyclic in its effects. The choice in the technology used to mediate interactions that people have with other actors in their network both *reflect* and *reaffirm* the nature of their relationship.

At the conclusion of their study, Licoppe and Smoreda point out a different concern that was revealed during their analyses.

Such analyses explore a relatively neglected scale in the socialization process, in between the large-scale perspective of social networks and the small-scale perspective of the micro-sociological construction of interactions.

This issue emerged from the necessity to study the user interactions that exist between the macro and micro worlds of analyses. This middle-scale represents a gap in methodologies that aim to connect analyses of different granularities. Improving research methods at this level can contribute to better relationships between macro and micro levels of study. This dissertation develops a framework that emphasizes a *meso-level* analysis, and applies it to understand how relationships and technology mediated interactions are connected, and in what ways social networks are "technologically embedded."

A second motivation for this research is the work of Donald Norman and the study of affordances (Norman, 1988, 1999). A concept that was invented by psychologist J. J. Gibson in (Gibson, 1977), affordances are the actionable relationships between an actor and an aspect of the environment that offers potentials for action to the actor. Gibson's affordances are relationships, and not just physical or visible properties of an object. Norman extends this idea to the human-computer interaction (HCI) field, and applies it to product designs, such as graphical user interfaces. He also distinguishes “perceived affordances” to describe what an end user perceives to be possible. The concept of affordances is especially important as new technologies continue to be developed. Our online activities have dramatically changed over the past decade, where Web 2.0 tools have given us an abundance of possible ways with which we can interact with each other. The pace of new technology development has not slowed, and research has struggled to keep up. A basic first step to understanding technological artifacts that mediate interactions is to study the affordances that each provide. Knowing how we use various tools can help inform us of what works well, what needs improvement, and what should be discarded.

As online interactions become more prevalent, many aspects of our offline world will begin to manifest online. Licoppe and Smoreda (2005) show that how we choose to interact with others is closely tied to our relationships with them. Our online social networks become more important, and thus our online interactions become important as well. The study of the relationship between the social and the technical means that we must give attention to the methods we choose in how we interact online; the affordances provided by various digital media will play an important role in helping to shape and define the relationships that we have. Therefore, it is not only necessary to study affordances for the sake of understanding digital media, but also for its effects on the relationships that are constantly formed and maintained. Norman has continued his efforts to refine our understanding of affordances (Norman, 2002, 2008), which I believe will help us to design systems that will greatly benefit our social interactions and relationships.

1.2 Research Questions

This research is exploratory in nature. I examine how technology mediates different interactions and develop a general framework to conduct future studies and answer the following research question.

- Research Question (RQ): How does technology mediate relationships that are formed in sociotechnical systems?

This is a broad question that can be broken down further. For the moment, I am defining *relationship* vaguely as a connection between two individuals, built by their interactions. I will cover the term in more detail later on. To answer the research question, I need to start by defining the interactions that are occurring inside a sociotechnical system, and develop abstractions that can connect it to relationships.

- RQa: What interactions take place in sociotechnical systems?
- RQb: What kinds of relationships are constituted by these interactions?
- RQc: How does technology enable the relationship-constituting interactions identified above?

RQa, RQb, and RQc require a framework that can connect interactions, relationships, and technological media. As with relationships, for now, I am using the term *interaction* loosely to describe any form of activity in a system where a member encounters any other member or a member-created artifact. Once I have established how the concepts are related and built different levels of descriptions, I can use the framework to help further my analysis.

Some aspects of the research question drive research in a parallel path. Understanding how relationships are connected to the technology that mediates them can help inform us how to use certain technologies to provide the most potential for developing strong and sustainable relationships. While not a separate research question, several issues will also be explored and discussed from the standpoints of the practical implications of the research. These include:

- How can we increase awareness of – and encounters with – other members?
- How can we support and encourage interactions based on these encounters?
- How can we increase connectivity across boundaries in the community?

As the questions indicate, one important topic of interest is the study of boundary spanners. These are members of a network that belong to two or more sub-networks that are otherwise unconnected (Aldrich & Herker, 1977). Boundary spanners serve an important role in the networks they reside in, and warrant closer study on how their interactions are connected to the relationships they form. There is significant research in boundary spanners and their activities (reviewed in later sections) that will help provide examples to ground this study when answering the research questions listed earlier.

This research will also help design strategies that can best support various user interactions that lead to relationships. Essentially, we are asking, "Who is out there? What do they have to offer? How can they interact?" There is an assumption made that interactions leading to relationships have a positive effect on both the users and the network as a whole. This requires viewing how system design is connected to technology mediated social capital. That concept can be extended as *sociotechnical capital* (Resnick, 2002). As sociotechnical capital resides in the relationship between technology and the user, increasing the amount of incidents where technology can mediate relationships between users will also increase the potential for beneficial actions to occur.

1.3 Preview of research

The goal of this research is to study how technology mediates the relationships formed in sociotechnical systems. New forms of online digital media offer different ways through which people can interact, demanding a need to understand how the media affects relationships. The affordances of the media might play a strong role in these interactions, and we need to study both the mediating effects and the types of relationships that people have. This research identifies gaps in the literature that have been used to study how people interact; many traditional theories and forms of study cannot keep pace with the rapid development that the Internet has experienced. I offer a new perspective of how to apply techniques such as social network analysis within the context of digital media. We can extend the traditional view of social ties to technologically mediated ties. While many aspects of face-to-face interaction can be extrapolated to the online world, we must accept that online interactions require new approaches to understand what is happening. I offer several studies that give examples of how artifacts with varying affordances might affect how members of an online community navigate within the system, and interact with other members. I explore a community of educational professionals and collect data on all artifact-mediated interactions between members over a two-year period of study. By examining these *pair-wise associations*, I am able to classify groups of user-pairs to make connections between the relationships users have with each other and the technology mediating the interactions. Closer detailed studies of some of the mediated interactions help ground my interpretations of the results.

Research in this area has also neglected to give focus to the study of *meso-level* analysis, the world of events between the forest and the trees. I analyze the online community under a meso-granular lens to demonstrate the potential of studying mediated actions, and how we can use such data to connect macro-level relationships to micro-level events.

2. Previous work

There are many different disciplines that apply to the research. Below, I will first cover social network analysis (SNA), which provides the methods to study relationships between people. Several topics related to SNA will also be examined, including dynamic social network analysis (DSNA) and boundary spanning. These topics are addressed as extensions to help guide my study. Importantly, I will examine how ties are used in SNA, and later present a different conceptualization that will drive the research. I also present a short review of social capital, followed by its application within technical systems, sociotechnical capital. These ideas are connected to how value can be associated with relationships in a sociotechnical system. I will also review actor-network theory, which supplements and expands my views on social networks, and exploratory sequential data analysis, acting as a guide to develop a framework to support my data analysis. Finally, I review several of my own works that explore some of the concepts introduced by the literature review.

2.1. Social network analysis (SNA)

Social network analysis is an analytic approach that examines people linked together by different relationships. Covered comprehensively in (Wasserman & Faust, 1994), SNA is a combination of theories, methods, and measurements that can be used to study the social structure created by relationships between people. Its focus is not on the individual or any specific personal attributes, but on networks as a whole.

SNA can be visualized graphically by plotting actors as nodes and relationships as edges, also called ties, in a network graph, or *sociogram*. In some cases, researchers have used SNA visualizations for studies not related to people; for example, SNA tools can be used to study the flow of currency between countries, where nodes are countries and edges represent the amount of currency being exchanged. Generally, when studying people, ties signify a relationship between particular actors in the network, whether it is formal or informal, friend or acquaintance, etc. In its simplest form, a tie's existence only requires that it be connected to two distinct nodes, forming a dyad (with exceptions for studies that apply self-directed edges). Ties are an important unit of study in SNA because they help determine the structure of the entire network. Many of the metrics used in SNA are related to ties: degree, path length, betweenness, etc., are all tie-related measurements or descriptions.

The sociogram graphical representation is a powerful analysis tool, able to help researchers identify points of interest such as clusters, boundary spanners, central and peripheral layouts, and other structural properties that otherwise would not be obvious in numerical data. SNA can be used for various kinds of analyses, across many different disciplines. Cross and colleagues explored many of these concepts in business organizational networks. They found surprising results regarding what they referred to as "bottlenecks" – those actors whose role in a network was to be connected to many different people that were not connected to each other. The study suggested that removing them caused networks to become weaker as a whole, and that the best way to alleviate bottlenecks was to distribute workload in a more decentralized manner (Cross, Parker, & Borgatti, 2000). Shortly after, they also found different ways to optimize a social network to become more effective in an organization (Cross & Parker, 2004). Some beneficial actions included fixing critical disconnects to prevent networks from dividing into sub-networks, and forming bridges between distinct networks to create new relationships and share information. Related to that is the study of structural holes (Burt, 1995). Burt found that brokering between unconnected actors benefits the broker by providing him with strong leverage, but closure of the gap by creating a tie is critical to realizing the most value (Burt, 2001). Another study focused on informal networks within a company, which had been largely ignored (Cross & Prusak, 2002). The study distinguished several roles by examining the network graph. Important roles included central connectors that had the most ties and knew who to find for information, boundary spanners that existed on the peripheries of their sub-networks and could connect different ones together, and information brokers that maintain communication between network sub-groups. The ability to identify roles based on where an actor resides in a network demonstrates the power of SNA in discovering information hidden within the structure of a network.

SNA can be applied in many other research areas. There are several that are of particular importance to this research, especially within online communities. Wellman and Gulia (1998) studied different types of online relationships and how they compare with those offline, exploring the possibility of intimate online relationships, and reviewing other potential extensions of offline interactions that the virtual world might be able to provide. Kollock (1998) observed online interactions where social action is taken independently of cost or benefit, finding particular incentive structures that help motivate cooperation. Garton, Haythornthwaite, and Wellman (1997) applied SNA in a computer-mediated communication (CMC) study and found evidence that CMC can affect the structure of a social network.

Today, there are online communities surrounding almost every conceivable topic, so it is no surprise that SNA has become more popular in research. Growing in parallel with SNA is the availability of different software tools. The development of SNA software has also fed back into driving SNA research, as increased computing power contributes to fast complex calculations and supporting large-scale network analyses (e.g. visualizing million node networks). Researchers can conduct studies based on network structures, and many of the calculations and measurements are made immediately available. Different forms of research can also be followed, and methodologies have been developed to work alongside certain software, such as exploratory analysis using Pajek (de Nooy, Mrvar, & Batagelj, 2005). Other software packages each have their own benefits, such as UCINet's[1] easy support of 2-mode

[1] https://sites.google.com/site/ucinetsoftware/home

data, or the statnet package built into the R[2] environment, offering great flexibility and statistical analyses.

Given the power of SNA, there are still gaps that have only recently started to be addressed. For example, sociograms are, by nature, static representations. They are snapshots of a network in a single moment of time, giving no hints as to how or why the network developed into a particular structure, or what it could potentially become. Examination of the evolution of social networks would be beneficial for research, especially in online communities, which can grow at tremendous speeds. I discuss some of the literature from that field in the next section.

There is also much that is unexplored in the analysis of individual ties in SNA. This is a curious gap in the area, as ties are the building blocks of the networks. While many studies measure ties at a macro-level, such as degree centrality or path length, there is less focus on the makeup of an individual tie. There is no universal standard in quantifying a tie, except to say that is strong or weak (to be discussed in future sections). Measuring the strength of a tie varies greatly depending on the context of the study.

2.2. Dynamic SNA (DSNA)

As an extension of SNA, DSNA attempts to fill the gap of not being able to conduct longitudinal SNA research. However, as DSNA is a relatively young field, there are few standards or methodologies. One path taken by Carley's research is to treat edges as probabilities, and use multi-agent systems to study network evolution (Carley, 2003). Carley redefined the traditional sociogram by adding probabilistic parameters into the edges, which can now be measured as the likelihood that it will form. Individual nodes were also given more emphasis; they are now treated as agents, and can potentially impact how a network will develop. Another approach is to explicitly utilize the time and order of social interactions to build the network (Berger-Wolf & Saia, 2006). These, and other frameworks, have built the foundation of DSNA. Research based on DSNA has started to apply some of these ideas. (Kossinets & Watts, 2006) analyzed the accessibility to bridges – members that connect multiple smaller networks – in a dynamic network with longitudinal data.

It is difficult to evaluate DNSA literature because the field is still young and growing. It is clear that longitudinal studies of social networks are important in our understanding of how networks evolve; the demand to study how social networks change over time will continue to push development in DSNA. The advantages are promising, and it has already been shown to be applicable in real world settings, such as terrorist networks (Carley, Dombroski, Tsvetovat, Reminga, & Kamneve, 2003). However, DSNA still requires more theoretical development and a more standardized approach before it can be used widely in more general situations.

[2] http://www.r-project.org/

2.3. Boundary Spanning

Boundary spanners are actors in a network that belong to two or more sub-networks that are otherwise unconnected (Aldrich & Herker, 1977). There are other terms that can be used synonymously, depending on the context. They have been referred to as bridges in SNA (Granovetter, 1973) or bottlenecks in some organizational research literature (Cross et al., 2000). Boundary spanners are important to networks because they are able to maintain effective relationships with a wide range of actors and help find the proper resources to solve problems (P. Williams, 2002). They tend to reside at network peripheries, and are usually not strongly connected to any particular network. However, they must be legitimate participants in all networks they belong to, have an understanding of the practices in each, and be able to negotiate relationships between them (Levina & Vaast, 2005).

Boundary spanners in online communities have become an important area of study; they are of particular importance to my current research (Joseph, Lid, & Suthers, 2007; Suthers, Chu, & Joseph, 2009). These studies have focused on the relationship between sociotechnical systems and boundary spanning activity. The advantage of boundary spanning in CMC is that it enables people to participate in more weakly tied relationships than face-to-face (Donath & Boyd, 2004). The role of boundary spanners can also be viewed in context of their network positions, where their value lies in how they extend other members' weak ties (Granovetter, 1973). This provides them with access to a greater source of new information. (Putnam, 2000) refers to these resources for potential action as bridging social capital.

A concept that brings together boundary spanning and sociotechnical systems is a boundary object.

Boundary objects are objects which are both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. (Star & Griesemer, 1989)

These artifacts are vital for boundary spanners to fill their roles. Research has suggested that the most effective boundary objects are tangible, concrete, accessible, and up-to-date (Carlile, 2002). Within sociotechnical systems, boundary objects are technical artifacts that are usually created or adapted by boundary spanners (Levina & Vaast, 2005).

2.4. Strength of ties

Social network analysis defines an edge between two nodes as a social tie (Wasserman & Faust, 1994). There are different approaches researchers take when trying to quantify the value of a tie. When using SNA outside of the social domain, there are no restrictions as to how a tie is measured, e.g. epidemiologists apply network analysis techniques to track disease spread and outbreaks. Strictly within the social domain, the most common measurement of a tie is its strength. Tie strength is generally understood as describing the level of closeness in a social relationship between two people, where strong ties consist of our close friends and family, while weak ties are made up of acquaintances.

There are several important studies that examine the strength of ties. Granovetter (1973) studied how weak ties can provide different advantages to actors in a network. His seminal research found that weak ties are more likely to serve as bridges between different network clusters. These ties give individuals access to networks they are not otherwise well connected to, and offer information and resources they could not obtain from their strong ties. Many of our close friends and family have redundant ties and overlapping networks, so there is a low probability of discovering new information through those means. Implicitly, weak ties help promote diffusion of information and innovation (Rogers, 2003). In his study, Granovetter measured tie strength based on amount of time in a relationship, emotional intensity, intimacy, and reciprocity of services. The subjects were students seeking employment after graduation. His results indicated that many of the participants were able to leverage their relationships with people they were connected via weak ties – as defined by his four measurements – in order to find post-graduate employment, supporting his argument regarding the value of weak ties.

Taking the opposite approach, Krackhardt (1992) addressed the influence and power of strong ties. His findings showed that strong ties based on trust and affection can help reduce resistance to change and provide comfort in uncertain situations. He observed a company's employees deciding on the issue of unionization. His results indicated that members who were connected via strong ties were more likely to offer emotional support and trust each other's opinions. Krackhardt measured tie strength based on frequency of interaction, level of affection, and history of interactions, measures that represented an alternate approach to that of Granovetter.

Both Granovetter and Krackhardt made valid claims regarding the importance of weak and strong ties. However, even though their arguments sound conflicting, they do not actually contradict each other. In fact, their studies have a fundamental difference because of each researcher's method in defining a tie. Granovetter measured tie strength based on the amount of time in a relationship, emotional intensity, intimacy, and reciprocity of services. Krackhardt's measurement was based on frequency of interaction, level of affection, and history of interactions. It is difficult to have a consensus on how to measure tie strength, but the two researchers are not alone in their difference of methods. Many other studies that examine tie strength are vague in what measurements were used or how it was obtained. Very few attempts have been made to unify the definition of tie strength by studying its indicators. Marsden & Campbell (1984) addressed the existence of the problem, noting that different researchers continue to use different indicators in tie strength studies. Additionally, they found that many of these indicators could be contaminated, e.g. frequency of contact might overvalue neighbors and co-workers, and diminish relationships with family members who live far away.

Many studies that apply tie strength do so with different indicators. Recent contact (Lin, Dayton, & Greenwald, 1978), organizational membership (Alba & Kadushin, 1976), and emotional support (Wellman & Wortley, 1990) are several examples of the different ways tie strength has been measured. There remains a gap in tie strength literature that directly addresses how to unify the many forms in which ties are considered strong or weak.

2.5. Social capital

Social capital is a concept with many definitions. Within the context of a social network, social capital can be described as the capacity to facilitate coordination and cooperation for a beneficial purpose. Inherent to the structure of relationships between actors, it is the potential for actions that can help support both individuals and the entire network as a whole (Coleman, 1988). Putman (2000) describes two forms of social capital, bridging and bonding. These concepts can be connected to SNA terms, where bridging is related to weak ties and bonding to strong ties.

Social capital is also a method of viewing value in a social network. Like traditional capital, it relies on the members of a community to contribute time and effort in order to grow. Unlike traditional currency, social capital is an abstract concept. It does not decrease when it is spent; instead, spending can actually increase the social capital in a network. As social capital builds, it potentially leads to increased interactions between actors in a social network (Tsai & Ghoshal, 1998). These interactions can directly influence resource exchange and serve as catalysts for knowledge and innovation. Because social capital is built upon interactions, this becomes a cycle in which a group's capital continues to build.

There is no universal standard to measure social capital. Different researchers have formulated many variations of how social capital can be quantified (Borgatti, Jones, & Everett, 1998; Burt, 2000; Lin, 1999). With respect to this research, I will not be focusing on the metrics of social capital; indeed, there are many other facets of social capital that can be examined, but this review only serves as an introduction to sociotechnical capital, which I will discuss below.

2.6. Sociotechnical capital (STC)

As more communities were born and began growing online, sociologists developed a need to adapt a method of study to accommodate the many affordances (Norman, 2002) of the online world. Studying the value of human interactions online naturally led researchers to apply theories of social capital. CMC introduced many new activities that were readily available to the masses of online users, but each contained its share of pitfalls. It became clear that social interactions did not occur in the same manner online as they did face-to-face (Hollan & Stornetta, 1992; Olsen & Olsen, 2000). A new approach in social capital would be needed for online studies (D. Williams, 2006). Williams approached the problem by developing a new scale, the Internet Social Capital Scale (ISCS) to measure bridging and bonding (Putnam, 2000) both online and offline. Ellison, Steinfeld, & Lampe (2007) also studied bridging and bonding, and how they are connected to social capital in the social network site Facebook. Their results suggested that Facebook users are able to use the site to help increase and maintain bridging social capital. Wellman et al. (2003)

noted that online interactions are forming new and different communities, and agree that they cannot be measured using standard indicators of social capital.

These, and other studies, do not give enough emphasis to the relationship between people and the technologies they are using. Any online use of social capital should necessarily include the technology that is mediating the interactions between users. Ties that are formed will be closely connected to the technologies that helped define them (Licoppe & Smoreda, 2005). In response to this gap in the literature, Resnick (2002) introduced the concept of *sociotechnical capital*. Where social capital describes the potential for action in a social network, sociotechnical capital expands upon that by embedding the capital with the affordances that technology now enables. For example, Galston (2000) noted that the lower cost of entry and exit for online communities can affect how well the community is kept together. Although Galston focused on social aspects (e.g. choice, affection, etc.), his results suggested that technology affords the user a much easier process to join a community online than offline. Whether it is a simple signup form or a complex questionnaire that requires validation, technology has a strong mediating effect on the types of interactions that occur online.

Just as social capital exists in the relationships between actors in a social network, sociotechnical capital resides in the relationships between users and technological artifacts in sociotechnical networks. Therefore, the design of such technologies is vital to maximize the potential for interactions that are beneficial to the network. Examples include simple navigation, easy access to resources, rich interactions with other users, etc. More complex interactions can also be studied through sociotechnical capital, e.g. why members identify with – and feel obligated to help – other members in the same network. Many suggestions from (Norman, 2002) still hold. Affordances of the technology offer online users the potential to conduct a particular action; sociotechnical capital is a pseudo-quantification of those affordances.

2.7. Actor-Network Theory (ANT)

Actor-network theory (ANT) is a sociological process of studying a network of actors, artifacts, and the associations that exist to connect them. (Latour, 2005) describes it as the "sociology of associations" and differentiates it from "sociology of the social." It views both human actors and non-human artifacts in the network under the same lens; the term *actant* is used to describe both such parts of the network. Actants are partly defined by the *associations* they have with each other. These associations do not only represent a relationship between actants, but also describe the history of that relationship. By tracing the formation of the association, one can learn how actants are navigating through the network and the relationship they have with other actants, both human and non-human. In describing the association between actants, you are implicitly studying both actants: by tracing human activity to understand how they interact with the resources and other humans, and tracing non-human activity to discover how the design of the artifact and network can influence the method by which actants interact.

Associations between actants have unique effects on how the network develops. The necessity to study the association between humans and technology is not new (Pinch & Bijker, 1984), but modern day sociotechnical systems offer many opportunities in this line of research. Applying ANT to a practical problem can help resolve complex associations between the social and the technical (Callon, 1986). Tracing associations provides a rich description of the relationship between actants, and how it was developed. The methods suggested by (Latour, 2005) can simplify an actor-network and provide meaning in the associations.

In *Reassembling the Social*, Latour describes five uncertainties that need to be accounted for. I describe the three that are most related to this research.

- Groups: to identify actors by group formation, which is constantly occurring
- Action: making connections between global and local phenomena, as there are numerous parties responsible for any single act
- Objects: to be treated as actors that also have a purpose

My line of research has a particular emphasis on how actors interact online, as mediated by digital artifacts in sociotechnical systems. Therefore, an analytic approach is needed that can view a network of actors connected via technological artifacts. Additionally, the pattern of interactions between actors is also important, and the need to trace that history becomes vital in viewing how relationships are formed and grow. I do not use ANT to discover why things happen in a given network, but rather, how. ANT can provide insight into several pieces of the process, especially through its concepts of actants and associations.

2.8. Exploratory sequential data analysis (ESDA)

Sifting through large quantities of data from social networks requires an understanding of how the network was formed and what relationships exist between the actors. ANT helps shed some light on how associations were built, but does not provide the tools for micro-analysis of the trails left by actors navigating through modern day sociotechnical systems. De Nooy et al. (2005) suggest exploratory methods of analyzing a social network using the Pajek software package. However, they do not link the research back to the technical systems from which the social networks were based on.

I find exploratory sequential data analysis (ESDA) to be well suited for the task (Sanderson & Fisher, 1994). It was developed as an empirical way of studying sequential data by integrating different observational data analysis techniques. Influenced by behavioral, cognitive, and social traditions, ESDA is a useful tool for human-computer interaction (HCI) and computer supported cooperative work (CSCW) researchers. It works best in situations where sequence is essential to the research and the software environment is supportive of the analyst's plan of research.

At the most general level, ESDA techniques involve a process of working from initial questions to final statements by manipulating observational data under the guidance of formal concepts. (Fisher & Sanderson, 1996)

In essence, ESDA only relies on observational data, questions concerning the data, and formal concepts that determine the types of operations needed. As its name suggests, the ESDA framework is meant for exploratory analysis, ideal for my needs. The research process is not linear, but rather, iterative, so that each cycle helps refine the methods being used. ESDA is not a confirmatory analysis method; it is likely that additional hypotheses can be generated during ESDA.

ESDA offers "eight C's" that will help guide researchers through their analysis. These C's are chunks, comments, codes, connections, comparisons, constraints, conversions, and computations. Briefly, their definitions are as follows (more detailed descriptions will be included later in the framework section for the specific operations that are used):

- *Chunks* are a combination of adjacent data elements. It is low-bandwidth filtering of unstructured events.
- *Comments* are notes, usually informal, that can be attached at any point in the analysis, to any data elements.
- *Codes* are abstract labels or keywords, coming from different possible sources, which help clarify data and provide a defined set of terms.
- *Connections* are relationships between data elements. They might not be linear, and the data elements do not need to be the same type.
- *Comparisons* examine the differences when data elements are handled differently or viewed under varying conditions.
- *Constraints* are filters to data that focus on a particular component of the analysis.
- *Conversions* reorganize data elements to uncover different patterns or results.
- *Computations* are summary representations of data.

There is no strict format with which to apply these "smoothing" operations. Which of the eight C's are used, or in what order, should be determined during the analysis. The flexibility offered allows the researcher the ability to organize the data as needed, and not waste time in unnecessary operations applied for the sake of formality.

2.9. Personal studies

Along with several members of my research group, two studies were conducted that begin to address some of the research questions. Our interest was specifically in regards to boundary spanning activities inside online communities. We began by analyzing the relationship between individual action and social affordances of the technology. This helped us develop a method of analysis for individual action.

Following that, we examined data from the same online communities to identify whether or not boundary spanning was occurring, and through what technological means. Below, I describe the two pilot studies and the contributions each one makes. However, detailed information is limited only to portions that are directly related to my research.

The research is based on the disCourse environment. DisCourse (http://discourse.ics.hawaii.edu) is an instance of a more general codebase, called Prometheus. The Prometheus software contains different resources that provide support for collaboration and community. The system homepage contains "stories" and resources that are publicly available without requiring a user account. This helps facilitate awareness of community events and topics of interest. Registered users have access to other user profiles, discussion forums, and workspaces. Profiles contain any details that a member chooses to submit, including email, pictures, contact information, and area of interest. Member contributions, such as stories or resources, are linked to the creator's profile, enabling members to find other people with similar interests. Threaded discussion forums are attached to a workspace, providing context for any particular topic. Workspaces are a "home" for any number of groups within disCourse. Inside, they house other artifacts, including resources, discussion forums, profiles of members in the group, etc. The main area of each workspace is a wiki page that can be edited by some or all of the workspace members, depending on the settings of that workspace. They are also paired with a discussion forum, allowing members to have dialogue about any general topics. Additional features of the system, and details about parts already mentioned, will be provided in later sections.

2.9.1. Transcendent Communities

An early conceptual paper (Joseph et al., 2007) looked at the boundaries that arose in online communities. These virtual walls, either through design or by accident, lead to silo-like structures that prevent different sub-groups from collaborating. For example, online learning environments are designed with similar structures to traditional classrooms. Just as classes are taught in separate classrooms, online classes generally have their own virtual space. There are practical reasons for these design decisions that make segregation necessary, e.g. maintain class identify, tuition restrictions on attendees, etc. However, these online silos inhibit different collaborative effects that can be conducive to learning (Derry & Fischer, 2005). Students and teachers participate in multiple nested and overlapping groups and are members of a larger *transcendent community*; yet this fact is not well supported by current online learning

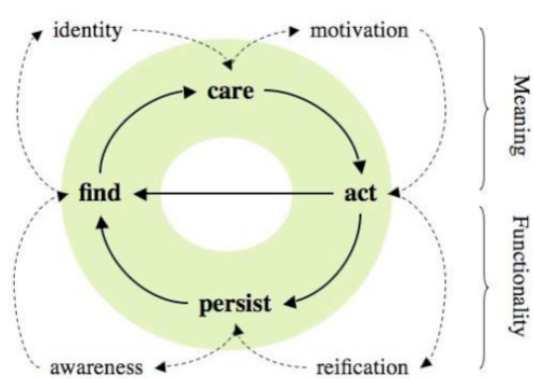

Figure 2.1: Activity cycle and related factors

environments. We examined some of the problems that silos present, and what efforts can be made to resolve them.

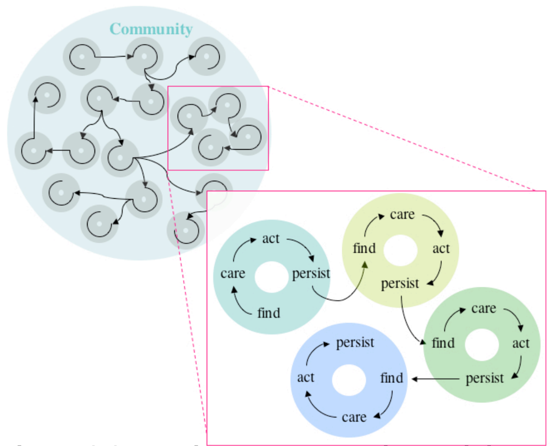

Figure 2.2: Persistence connecting activity cycles (image created by Viil Lid)

Any network-level or social phenomena exist only because of mediated acts in a given setting. We follow a process to "localize the global," as suggested by actor-network theory (Latour, 2005). In studying user activity within the sociotechnical system, we discover certain patterns of participation by many of the users. First, a user must find some trace of past action. This depends on the system making the user aware of different artifacts. Second, the question arises to whether or not the user cares about the artifact. This might depend on if the user identifies with the artifact, either through earlier interactions or if they both belong to a shared community. Third, the user must perform some action on the artifact. Last, the action must leave a persistent trace. This provides a lasting trace that can be found by other users, thus facilitating the restart of a new cycle. We define this pattern the find-care-act-persist cycle (FCAP). These actions can be understood as part of a participation-reification duality (Wenger, 1998), where acts of participation are intrinsically bound with reifications that leave traces of the act. Figure 2.1 presents a visual representation of the FCAP cycle. Figure 2.2 expands on the concept to show how different members of the community can repeat this cycle. In some cases, the cycle completes and offers additional reification for other members; in other cases, the cycle is incomplete and there is no contribution made back to the system. Continuous cycles will drive interactions in the community and create resources for other members to find.

We further expand FCAP activity to include interactions with members in other communities. Boundary spanning via individual navigation in conjunction with boundary objects can help create pathways through which members from two distinct communities can interact. Figure 2.3 is an example of how a user can leave traces of their activities in the FCAP cycle for another user to discover and use. Each circle denotes the FCAP activity cycle. They are grey if there is potential for FCAP, and pink if the full FCAP cycle has been fulfilled. By following the arrows, we are tracing an artifact through the system as it interacts with users. When the artifact (represented as small squares) has successfully helped connect a user cross boundaries to another community, it will change from grey to orange. We can also trace how the user is navigating through the system. The blue (in grayscale, dark) and green (light) arrows are two examples of boundary spanners as they traverse between communities by interacting with different boundary objects. In this example, the users are able to "discover" each other when they meet in a sub-domain connecting two different communities by following different artifacts. From a design point of view, this recommends the need to create opportunities for users to discover reifications left by other users. By understanding the activity cycle, we can determine which part of the process correlates with certain actions, and it can benefit from technical intervention.

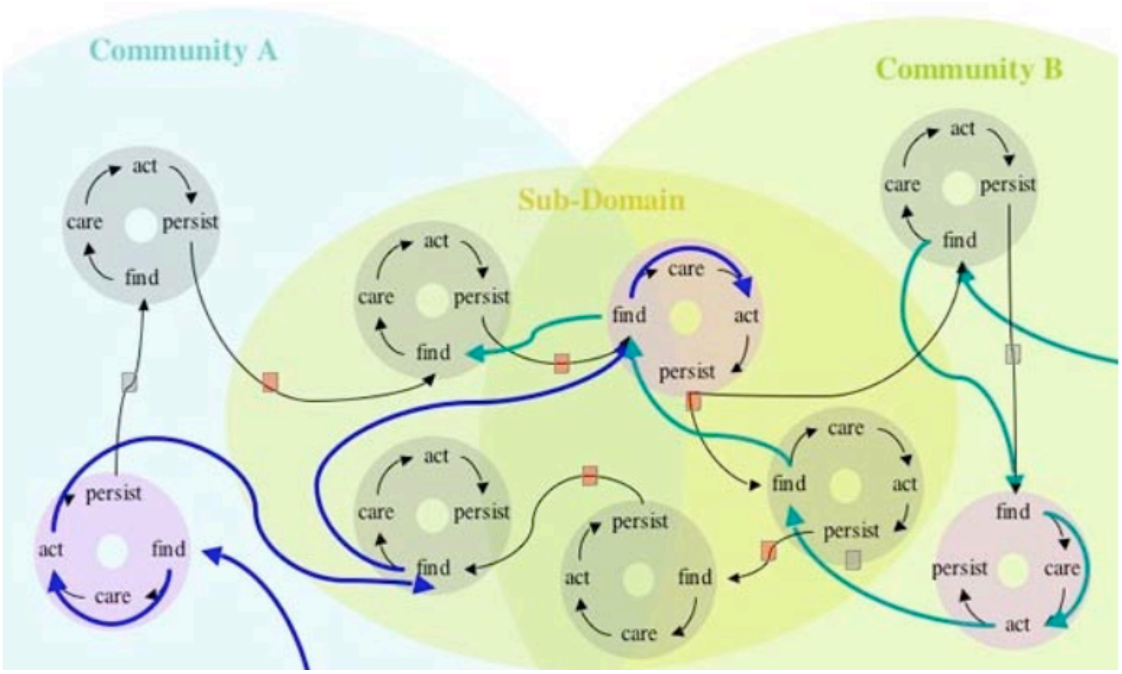

Figure 2.3: Pathways connecting communities (image created by Viil Lid)

2.9.2. Bridging Sociotechnical Capital in an Online Learning Environment

The previous study explored the idea of boundary spanning in sociotechnical systems. We developed a simple framework of the FCAP activity cycle that shows how a user navigates around the system, and potentially through virtual walls between communities. We follow up that analysis by examining data in disCourse to determine how, if any, boundary spanning activities were occurring.

The next study (Suthers et al., 2009) is also based on the online learning environment disCourse. We considered learners to be potential resources to each other who are prevented from leveraging this potential by the virtual silo structures that are in place for logistic reasons (e.g. classroom identity). We wanted to see whether members were discovering other members or resources that were outside of the context to which they were originally assigned.

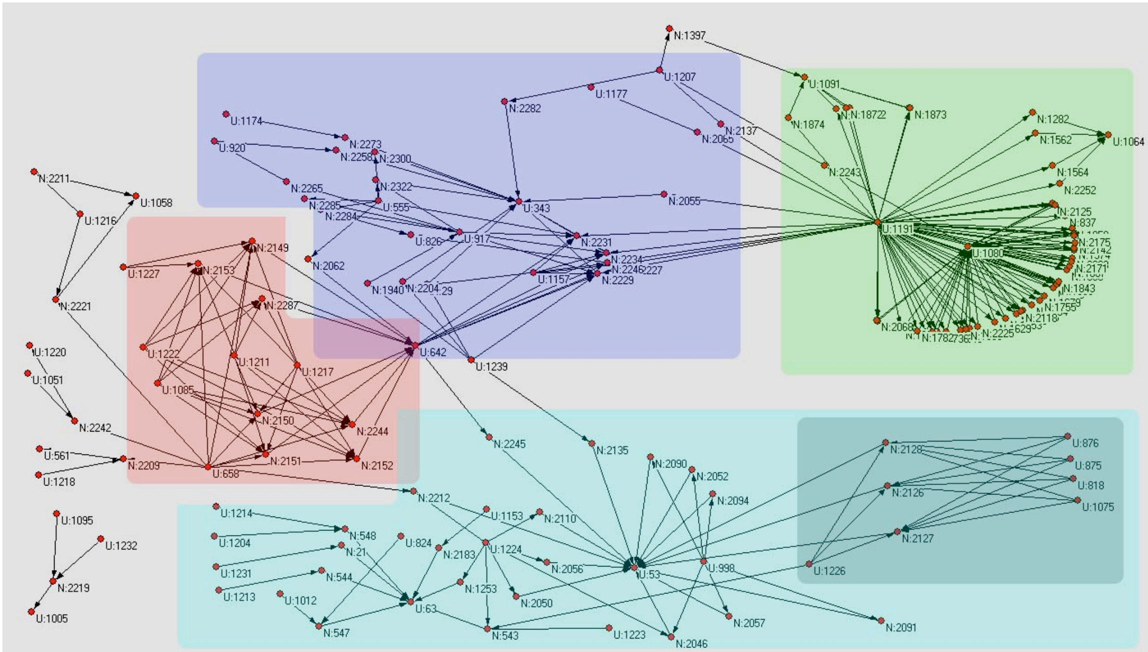

Figure 2.4: Fall 2007 wiki associations

The disCourse system has several different components that enable interactions to occur between members. There is a resource database that enables users to store links to files, URLs, etc. The site is divided into virtual workspaces that members can belong to. Examples of workspace uses include classes, research teams, or interest groups. Workspaces are also hierarchical and can contain sub-workspaces. Threaded discussion forums exist for members to compose messages about different topics. Wiki pages allow members to post and edit information in a shared space. Both discussions and wikis can exist inside a workspace.

Our study was based on the assumption that if person A accesses something (e.g. reads a wiki page) that has been created or modified by person B (e.g. created a wiki page), then person A has derived some value from the presence of person B. We focused on specific technical artifacts (discussions, wikis, resources, and member profiles) that members can directly access, and were created by another member. Because we wanted to find out whether bridging was occurring, we limited our analysis to events where someone accessed an artifact that was provided by another member who does not belong to the same group (i.e. workspace).

For the purposes of this study, we constructed custom network graphs using Pajek to display relationships between actors and artifacts. We followed several of the smoothing operations from ESDA to abstract our data from basic log files to create the network graphs. Log data was chunked and coded (Sanderson & Fisher, 1994) to form events; connecting procedures were used to identify the links between actors and artifacts. The network graphs were not traditional sociograms used in SNA, but rather a bipartite network consisting of two modes – actors and artifacts (more

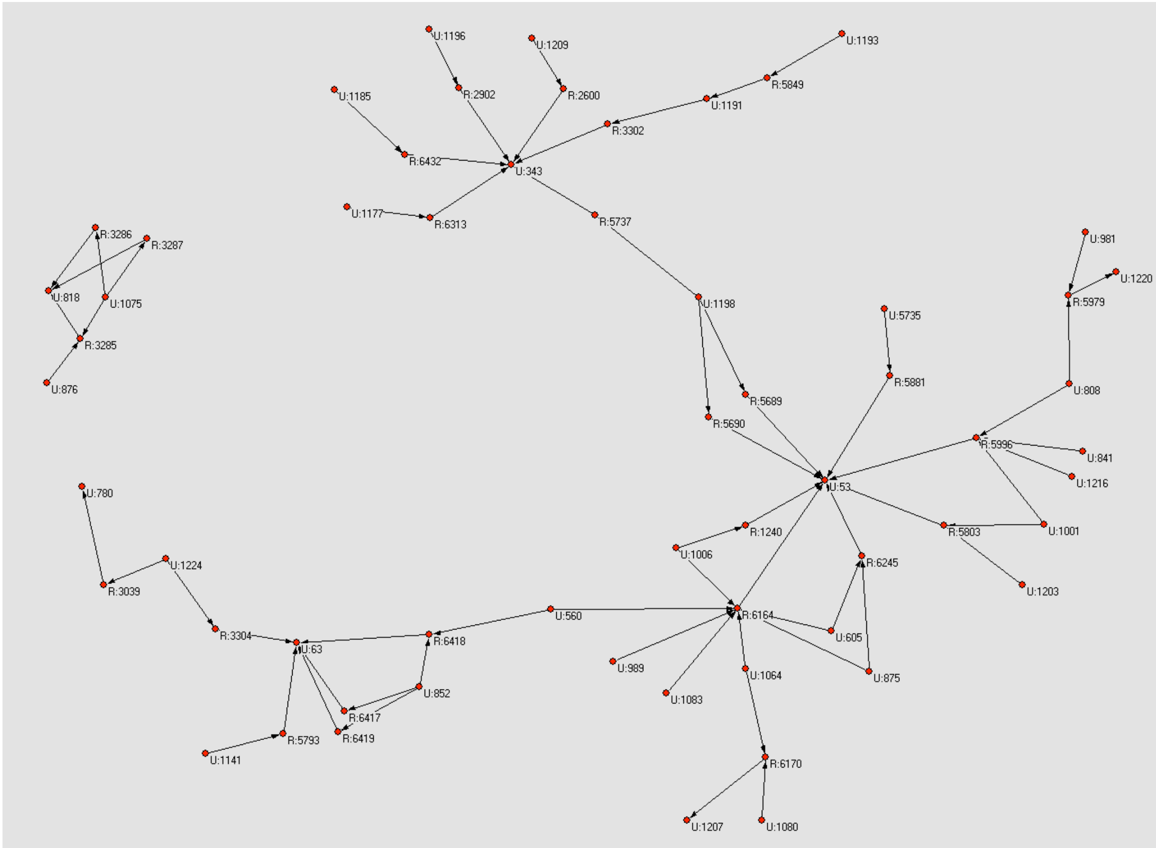

Figure 2.5: Fall 2007 resources associations

detailed information regarding this type of network graph will be discussed in future sections). Queries were constructed to trace user activity in two separate academic semesters. For each of the four artifact types, we generated a unique network graph for each semester. To account for students that resided in workspaces because of class requirements, we removed all ties that were created between members in the first two weeks of each semester.

Viewing Figure 2.4, it appears that there is a high amount of boundary spanning as mediated by wikis. The colors in the graph were manually added to highlight clusters that defined the virtual boundaries between different groups. However, upon a more detailed inspection, we found that much of the activity resulted from teachers that added students to sub-workspaces after our “two week rule.” Once we had adjusted for that issue, the network showed significantly less activity; wikis were not artifacts that mediated much boundary spanning activity.

Figure 2.5 is the graph of actors connected to resources in disCourse. Along with profiles (graph not shown), resources account for a large number of boundary spanning activities as compared to wikis and discussions. One reason why this might be the case is that profiles and resources are usually publically available for access. Discussions and wikis tend more to be concealed artifacts since they are associated with workspaces that are private (e.g. classes).

Our analysis was a small step in studying user activity in an online learning environment. We developed a different method of viewing how artifacts mediate relationships between users. Several simplifications were made, e.g. how we defined a tie, what members did with material they found, etc. One improvement could be to use more complex definitions of ties (e.g. round-trip interactions rather than one-way), potentially yielding better results. Nonetheless, the study suggested that bipartite network graphs have potential to be a useful analytic device for studying the relationship between actors and artifacts. The study also informed us on how to improve our definitions of interactions and relationships. We obtained a better understanding of how different properties of digital media can affect online boundary spanning. These results provide many of the details on how to build a better framework for this type of analysis.

The development of the FCAP activity cycle showed the need to study the traces that actors leave behind when they navigate through a sociotechnical system. The importance of discovering how they transcended their silo boundaries further drives the need to examine what role technical artifacts play. By following the study and viewing the differences in artifact-specific network structures, we begin to realize how affordances of mediating technologies can affect both network structure and the amount of interactions that are occurring. These two studies helped provide some of the fundamental framework for how I can study sociotechnical networks. However, they also revealed some of the deficiencies in current research methods.

3. New Framework

In this chapter, I address several of the gaps in the previously reviewed literature. I present new ideas on addressing some of the issues, a reconceptualization of social network ties, and a mediation model view of network data.

3.1. Addressing ties in social network analysis

This section is a summary of the research in (Rosen & Chu, 2011).

Tie measurement is an analytic foundation of social network analysis. Ties have been most commonly measured with respect to the concept of "strength" of the tie. The strong tie – weak tie dichotomy is conceptually misleading, and the science of networks lacks a clear conceptualization of ties that allows for consistent operationalization of the concept. This conceptual gap has led researchers to measure ties in a contextually specific manner, often unique to their particular research design. A reconceptualization of the tie concept as the *utility* of a tie is necessary. It should provide a multi-dimensional taxonomy that can be used to assess appropriate tie measures.

The comparisons between strong and weak ties can be conceptually ambiguous, as its use in the literature does not represent polar sides of a continuum; instead, they are two different theoretical areas that operate independently, with strength representing socio-emotional closeness and weakness representing access to resources. A large number of measurements have been employed in a myriad of contexts, yet the science of networks lacks clear conceptualization of ties that allows for consistent operationalization of the measure.

Contrasting views relating to the strength of ties can often come from different indicators. There is an abundance of research that studies and applies the strength of ties (Friedkin, 1980; Granovetter, 1973; Wellman, 1982), but little research that directly addresses tie strength indicators or how they are interconnected with each other, and the work that exists is often vague in exactly what measurement was being used or how it was obtained. Those studies that are specific to a particular component are not always compatible with the different social aspects of a tie, and are difficult to generalize outside of the study.

The large number of dimensions that can encapsulate tie strength present it as a misleading term because of its multi-dimensional nature, and attempts to measure it need to be categorized based on the indicators used. Tie utility can be used as a description of the combination of various indicators used to describe a tie.

3.1.1. Issues with a Uni-Dimension Tie

If one draws conceptual conclusions based on the topics of Granovetter (1973) and Krackhardt's (1992) research, it risks the assumption that the two researchers were studying two opposing measures of the same concept (even their paper titles support that idea: "The Strength of Weak ties" by Granovetter and "The Strength of Strong Ties" by Krackhardt). However, their approaches to studying ties were based on different indicators. Although both make valid claims regarding the strength of strong and weak ties, they used different methods to measure tie strength. That is, Granovetter's weak ties are not the same as Krackhardt's weak ties; similarly, Krackhardt's strong ties are not the same as Granovetter's strong ties. Both studies make sound arguments concerning the effects of strong and weak ties, but only with respect to each researcher's definition. Consider that one of Granovetter's measures is time in a relationship. This indicator will identify many new acquaintances as weak ties, although it will ignore the types of interactions between those people. Contrast that to Krackhardt using frequency of interaction as an indicator in his study; while two people having many interactions in a short amount of time might suggest they have a strong tie, it ignores how long they have maintained this level of companionship.

There is an abundance of research that studies the strength of ties (Marsden & Campbell, 1984), but few that directly address concerns regarding tie strength indicators or how they are interconnected with each other. Many studies that examine tie strength can be vague in exactly what measurement was being used or how the data was obtained. The studies that are specific to a particular component (Shi, Adamic, & Strauss, 2007) are not always compatible with the different social aspects in a tie, and are difficult to generalize outside of the study.

Marsden's research (Marsden & Campbell, 1984) begins to address the problem of tie strength having multiple descriptions, although it is limited to only distinguishing time spent in a relationship and depth of relationship. Additionally, the research was aimed at unifying tie strength through a single indicator (according to Marsden, closeness) rather than using multiple indicators as different measurements. However, many of their findings have helped inform us in the direction we have taken in our theory, including the different effects of indicators and predictors, and contaminations that can occur. Petroczi and colleagues (Petroczi, Nepusz, & Bazso, 2007) conducted a study of tie strength in virtual communities. Their research in developing the Virtual Tie-Strength Scale (VTS-Scale) has been one of the few efforts to provide a quantitative measure of tie strength, expanding on Marsden's original list of indicators. Their study offers empirical data that demonstrates the asymmetric nature of ties, and the VTS-Scale shows promise in steering the description of a tie away from solely being either strong or weak.

Table 3.1: Different approaches for tie strength measurement

Tie strength measurement	*Research approach used*	*Source*
Frequency of contact	Often v. occasionally v. rarely	(Granovetter, 1973)
Mutual acknowledgement	Knowledge of another member's research (binary and not necessarily symmetric)	(Friedkin, 1980)
Duration of contact	**---**	(Wellman, 1982)
Social homogeneity	Duncan's socioeconomic index (SEI)	(Lin, Vaughn, &

		Ensel, 1981)
Overlap of organizational membership	Questionnaire of frequent intellectual discussions and topics, intellectual groups, and correlations against influences regarding Vietnam	(Alba & Kadushin, 1976)
Recent contact	Small world experiment, chaining letters that asked about last physical contact	(Lin et al., 1978)
Affection	Questionnaire assigning “friend” to a symmetrical relationship	(Krackhardt, 1992)
Closeness	---	(Marsden & Campbell, 1984)
Emotional support	Surveyed emotional aid (minor, family advice, major)	(Wellman & Wortley, 1990)
Services rendered	Surveyed services (minor, lending/giving, household services, organizational dealings)	(Wellman & Wortley, 1990)
Public acknowledgement	Facebook: friend, picture friend, roommate	(Lewis, Kaufman, Gonzalez, Wimmer, & Christakis, 2008)
Redundancy	Triadic closure	(Shi et al., 2007)
Accessibility	Structural holes	(Burt, 2001)
Affiliation	Shared membership in a group.	(Breiger, 1974)
Exchange	Amount of exchange between nodes	(Barnett, 2001)

We examined a sample of tie studies in the literature that measures tie strength using different measurements. Table 3.1 shows some of the examples of how researchers studying tie strength have varied in their particular approach. While comparing Granovetter and Krackhardt’s seminal works is the clearest example, there are many other cases of how tie strength is studied in different ways. Each study also differs in how information regarding ties is collected, using surveys, scaled questionnaires, binary values, online user data, etc. The use of so many different measures and data collection methods makes direct comparisons difficult.

We do not presume Table 3.1 to be a comprehensive list, as there are other studies that apply additional approaches to the measurement of tie strength. As social network analysis continues to evolve, especially with regard to online social networks, additional measurement techniques should be expected. It is unfortunate that most of these studies cannot be directly compared against one another; there is no single measurement that would allow such comparisons. Although in many cases, a direct comparison would not be practical or valuable, several studies that apply similar measures could be compared and contrasted to better understand how they might relate.

3.1.2. Impact of a Uni-dimensional Tie

We suggest that tie strength cannot be measured as a single general value. Measurements used in literature have varied greatly from different studies, yet all claim to measure a single term. This conflict extends beyond research studies; social network analysis software tools such as Pajek or UCINet offer few courses of action to analyze the individual edge that identifies a tie. There are options that enable users to input the weight of an edge, which can be visually represented by a thicker or different colored line, but that is often the limit of the level of tie analysis available. Cyclically, the lack of any visual analysis tool with a deep focus on ties might further contribute to the parallel shortage of detailed studies regarding tie strength.

3.1.2.1. Explicit Ties in Online Communities

Tie strength research has continued to evolve over the past decade. The popularity of online social networking sites has opened up a plethora of opportunities to examine network ties. For example, Lewis (2008) used a large dataset from the social network site Facebook to analyze network ties. In the study, tie strength was measured by seeing whether a person had "friended" someone else on Facebook or had pictures of the other person, whether the other person had pictures of them or lived in the same (real world) housing group. This type of data would not have been available prior to the expansion of online communities such as Facebook or MySpace; additionally, it provides a clear example that the study of ties is not only connected to the development of online communities, but also still continuing to grow and evolve.

Social network sites usually have explicit ties that are created when a friend request is made from one person to another. Whether in Myspace, Facebook, LinkedIn, etc., each of these sites have functionality that allows a user to make a friend request, and conversely, accept or decline friend requests that are made to them. If accepted, these create an explicit tie between two members. In conducting any type of network analysis within one of these sites, it is relatively straightforward to consider a completed friend request to be a tie between two members. As simple as this sounds, it nonetheless presents several problems. Primarily, it is difficult to compare ties across different sites. Is a Facebook friend the same as a Google+ friend? Can they be compared against each other? Clearly there is no requirement that any friend added in one social network site must be added in any other site. In fact, it is likely that people have different sets of friends across the various social network sites. Each site typically has a different focus from which it was built, whether it is for professional connections in Linked In or people that have musical affiliations (musicians, fans, producers, etc.) in Myspace. Without any way to describe and compare how friends might differ across different sites, it becomes very difficult to understand what "friending" someone actually means.

3.1.2.2. Weighing Ties

Another problem is that these ties have no form of weight. Clearly, a member's Facebook friends are not all equal. Some might be close family members, while others might be people they have never met. Being able to put specific friends into different groups, or the Circles feature in Google+, helps alleviate this issue to a degree. However, the primary function of a group feature is not necessarily to sort by level of friendship, but rather a feature to help control privacy issues; each group or circle has different levels of privacy that allow members to specify what artifacts that group can and cannot see. Therefore, groups do not solve the issue of distinguishing different ties between people; a person might have familial ties with their parents and siblings, but might not always share the same online content with both groups.

3.1.2.3. The Need for Multiple Dimensions

Unfortunately, social network analysis in online communities has been applied without any resolution to the problem associated with using a single dimension for ties. Just as SNA research continues using tie strength as a single concept to describe any number of tie measurements, the same problem manifests itself in online communities. Any "friend" link between two members in a social network site reveals very little information about what the relationship between those members actually means. Are they family? Friends? Acquaintances? Co-workers? A great deal of information is needed that can give some context to how two people are really associated with each other.

3.1.3. The Utility of Multi-Dimensional Ties

Multi-dimensional ties offer a more accurate understanding of how the study of ties can be framed. Research from (Rosen & Chu, 2011) suggests that tie measurements can be classified into three dimensions: socio-emotional closeness (e.g. social homogeneity, affection), resource potential (e.g. overlap of organizational membership, service rendered), and accessibility (e.g. frequency of contact, redundancy). These dimensions provide relationships between the various measurements that are currently used to describe ties, and an improved understanding of how the indicators can be linked. Rosen & Chu's reconceptualization has implications in understanding the connections in the social aspects of relationships between actors, some of which go beyond the scope of this research. However, the process of approaching ties as a meta-object that is built on various components is an important concept that I plan to leverage.

For practical application, a multi-dimensional tie approach also has advantages when studying relationships with regards to digital media. In the Web 2.0 world, there is an abundance of methods by which people can interact with each other. A one-dimensional tie is both limiting and misleading. A classic social network tie denotes a type of social relationship between two actors (Wasserman & Faust, 1994). However, a tie between two actors based on online interactions has a different meaning. There does not necessarily need to be an expectation of any social relationship at all. Online interactions have opened up a new view of how traditional social network ties can be viewed. In many cases, people who interact online have never met face-to-face. This begs the question: how can there exist a social relationship between people who have never met? Therefore, an online tie must represent a different form of relationship rather than a typical social tie.

If social ties are based on interactions, online ties can also be represented in a similar manner. Social tie measures include frequency of contact, overlap of organizational membership, recent contact, etc. These measurements give different descriptions of how people might interact differently with each other. We can extend that to online interactions by including data on the types of digital artifacts that mediate such interactions. Rather than apply a "traditional tie" between two given actors, we can better describe their relationship using aggregated data on *how* they interact. This would be an inherent feature of using a multi-dimensional tie approach, as it would require that the description of any relationship be based on many different kinds of interactions. The idea that associations can be distributed across multiple media types (Licoppe & Smoreda, 2005) entails the nature of a tie to be encompassed by more than a single dimension.

3.2. Online Log Data

Online interactions offer a major advantage when comparing against offline interactions – they can be completely recorded. When users interact with each other online via any website, their actions are usually logged on the server that is hosting the content. There are different levels of logging services, which will vary based on how much information is stored. For example, the popular Apache web server offers many logging features, capturing security issues, errors, and all requests that the server processes. From the perspective of the end-user, it can record every link that is clicked or followed, any information that is entered in a form, and generally any user action that results in communication with the server. There are other applications that can store even more data, such as all mouse movements. These and other logging methods offer a new glimpse into user actions that might not be readily available when studying SNA offline, which can rely heavily on surveys and questionnaires.

The ability to have access to all user activity that is conducted in an online system affords an opportunity to study online social interactions in a different manner. We can identify every possible method through which people are interacting online, whether it be emails, synchronous video or text chats, asynchronous interactions such as discussion forums, etc. Different companies and groups are also continuing to develop new methods of interaction, such as Facebook's wall, various picture storing and tagging sites, short Tweets, instant messaging systems, etc. In fact, those systems are examples of interaction methods in different communities; within the same community, there are still multiple methods of interaction. Twitter has popularized micro-blogging tweets, but also allow users to link images and send private messages. Facebook offers even more ways, with a public wall, private messages, photo tagging and annotation, and "liking" different events and objects.

3.3. The Distribution of Ties Across Multiple Media

Applying multi-dimensional ties in online communities brings a different view of how ties might be constructed. Online interactions are logged and can reveal exactly what users were doing in any system. Even within a single sociotechnical network, users are offered many different ways through which they can interact with each other. Each available method requires some digital artifact that mediates the interaction and provides different affordances. For example, a user might choose to send a private text-based message to another user. There will be some limit to the message size, and they can assume there will be some time delay before the message is read. They can also expect some level of privacy, as the message should only be sent to another individual. However, the same user could choose to send the message via a public chat room where the recipient of the message is also present. This method has a pre-requisite that both users are co-present, but guarantees that the message will be read almost immediately. However, there might be a different limit to the size of the message that can be sent; additionally, since the message was conveyed in a public chat room, there could also be other users in the same virtual room that have access to the message as well.

With many different media options available, each with its own set of affordances, online users have the luxury to choose their method of interaction. These mediating artifacts provide different boundaries on how someone can interact with others. Under a multi-dimensional tie approach of studying user interactions, we can base an online social network tie on the aggregated interaction methods used by any pair of users. That is, we can reconstruct any association between two online sociotechnical system users by the total set of interactions they have participated in via different mediating artifacts.

Interactions can be spread across different locations, spaces, and time. Different modes of communication present multiple temporal granularities. The affordances of different media bridge many of these variables. Depending on the choice of media, a person can connect across geographic space, virtual space, or even choose the effect that time can have in how they interact. Rather than trying to summarize all the different kinds of possible interactions into a single tie dimension, we should respect the differences that each media type affords to user interactions, and describe and compare them accordingly. A multi-dimensional tie framework is better suited in an environment where there are different possibilities for people to choose from in their mode of interaction.

3.4. Associograms

I now introduce a new way to model how people can be associated via mediated interactions, based on findings in my research group and expanded in (Suthers & Rosen, 2011).

3.4.1. Actor-network theory influence

Working with other members of my research group, we are inspired by several of the concepts used in actor-network theory. When studying online interactions, there are significant differences in the relationships that people form when compared against offline relationships. We follow a simplified version of Latour's concept that a social system is comprised of mediated associations (Latour, 2005). This understanding better describes online interactions between actors that might not have ever met in person, or even had direct interactions with each other. Some form of digital artifact is necessary to mediate online interactions, which can be aggregated to describe the association between two actors. Just as social ties are assembled to create a sociogram, we can build a network based on mediated associations. We call this network an *associogram*. The term is influenced by the sociogram model. However, we wish to avoid many of the socio-emotional connotations of the terms "relationship" or "tie" and feel there is a stronger parallel with Latour's mediated associations.

3.4.2. Description

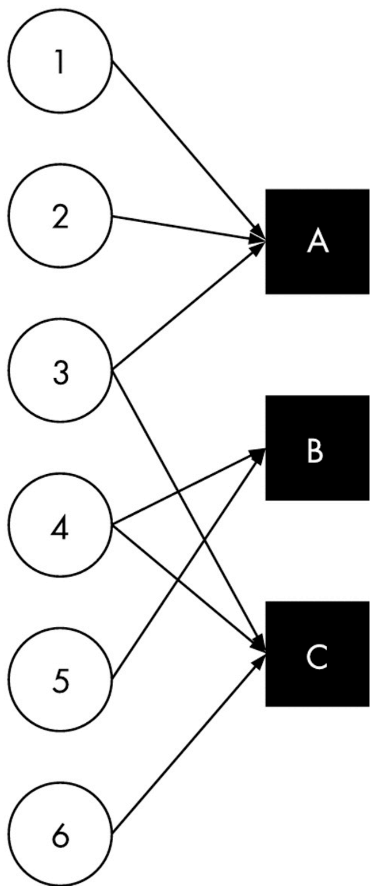

Figure 3.1: Example of bipartite network

An associogram is a type of bipartite network. Two nodes in a bipartite network are affiliated if they participate in the same group or event (Breiger, 1974). Bipartie networks are also called affiliation networks, as they represent how actors are affiliated with a set of groups or events. The major difference between an ordinary network and a bipartite network is that there are two types of nodes in the latter. Ties between nodes can only connect nodes of different types; no two nodes of the same type will ever be connected. For example, a bipartite network can be used to represent organizational membership. The two types of nodes would consist of actors and organizations, and ties would represent membership. Therefore, any actor node could have ties connecting it to different organization nodes. Conversely, every organization node would be connected to actors who are members. Figure 3.1 is an example of a bipartite graph, where the set of numeric circles represent one type of node, and the alphabetical squares represent another. Note how edges only connect circles to squares.

There are many practical uses that can leverage the power of bipartite networks. Examples include co-author citation networks (Chen, 1999) and social roles (Davis, Gardner, & Gardner, 1941). Existing SNA techniques and measurements can be applied to bipartite graphs; degree centrality, betweenness, path length, and many other metrics are calculated in the same manner.

Within an associogram, the two types of nodes are actors and artifacts, although both can be considered ANT-ian actants. One difference between associograms and bipartite networks such as affiliation networks is that associograms can further divide the node types, for example, into various types of artifacts. Another difference is that bipartite networks are usually treated as undirected graphs, while associograms are directed graphs: the edges that connect an actor to an artifact can be directional. The direction of the arrow is used similar to that of a state dependency diagram. An edge pointing from an actor to an artifact can be understood as "the state of this actor is dependent on the artifact." A simpler description is that the actor has accessed the artifact in some manner, e.g. downloaded a file or read a message. Conversely, an edge pointing from an artifact to an actor is read as "the state of this artifact is dependent on the actor," or that the artifact was created or changed by the actor.

The difference between traditional sociograms and associograms can be clarified by translating between the two types of network representations. A sociogram tie represents the relationship between two actors. That relationship is built upon many different factors. If we reconceptualize the tie as a mediated association, it can be understood as the aggregate of all the mediated interactions that have taken place between the actors. By unpacking these interactions, we make the artifacts that were responsible for the mediating actions explicit. These artifacts are then represented as a second set of nodes in the associogram. Each actor-actor dyad in a sociogram is transformed into a set of actor-artifact-actor triplets. A single tie is unpacked into multiple associations that mediate the relationship between two actors.

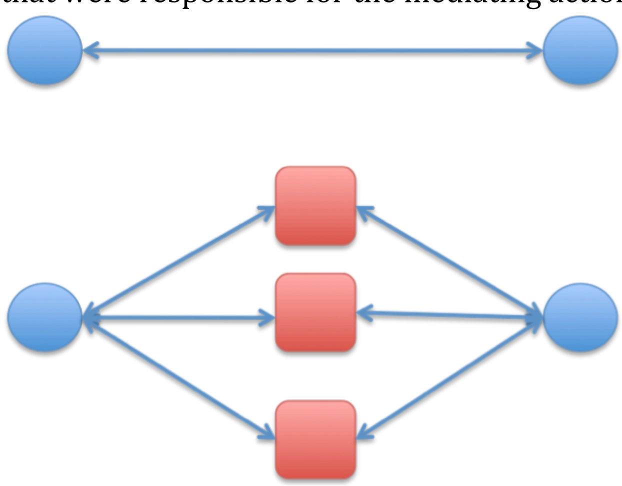

Figure 3.2: Comparing a sociogram tie (top) with associogram ties (below). Circles are actors, squares are artifacts.

The process can also be completed in reverse so that the data from an associogram can be reduced to form a sociogram. Given two actors who are connected by any number of mediating artifacts, all the associations between them can be compressed to form a single sociogram tie. To mitigate loss of data from compressing the associations, information regarding the artifacts can be represented as the strength of the tie between the actors, although not all variables can be easily captured. In Figure 3.2, the top portion represents a traditional sociogram tie between two actors (the circular nodes). The bottom portion represents what might occur if we expanded the tie, showing three artifacts (the square nodes) that are mediating the tie between the two actors.

Associograms are not limited to any single form of media. Although one set of nodes represents non-human artifacts, it is not limited to any single type of artifact, even in the same associogram. This offers an opportunity to study interactions through a multi-media view. We can observe networks where actors interact with each other through any number of methods, with the ability to compare the effects that different artifacts might have on network structure, dyadic relationships, boundary spanning, etc.

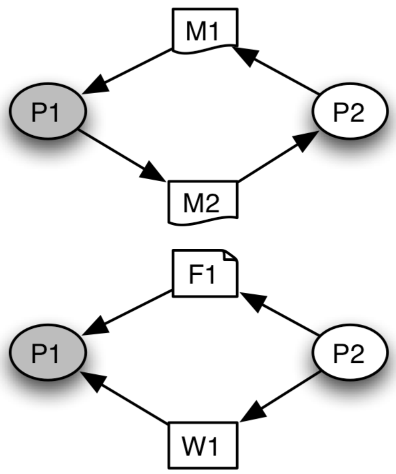

Figure 3.3: Different mediated associations. P = Person, M = Message, W = Wiki, F = File

The associogram provides information on how actors are connected to each other. The interactions between them take place through some mediating artifact. Analysis at this level provides a *mediation model* from which we can view a given network of actors. If we choose to view only two actors, we have information on both what artifacts mediate the associations between the actors, and also the pattern of interactions based on the directions of the associations. These patterns can help identify the type of relationship that the actors have with each other. For example, Figure 3.3 shows two examples of how actors might interact. In the top image, the actors interact via messages with each other. The direction of associations is cyclic, suggesting that the actors are having a balanced conversation. In the bottom image, the actors interact via a wiki and a file. However, the direction of the edges both point toward one actor, suggesting an unbalanced producer/consumer type of relationship. Other patterns of interaction, when combining multiple artifacts and directions, can be classified.

3.5. Entity Event Contingency Model

I would like to incorporate the associogram into a framework that can connect it to other representations as well. A process was detailed in how we can shift between sociograms and associograms, but there is a need to analyze associations, in of itself, in finer granularity. In the same manner by which we unpack a social tie into its constituent mediating associations, we can further examine the pattern of interactions that define how an actor and artifact are associated. For that level of analysis, we need to know what interactions are taking place, the sequence of events that correspond with the interactions, and the entities involved in any given event.

The Entity-Event-Contingency (EEC) framework offers the tools and guidelines that can support this form of research (Suthers, Dwyer, Medina, & Vatrapu, 2010). The EEC framework is a conjunction of three different representations: entities, and their relationships with each other, are the domain model and capture all actors and artifacts in the system; events, which are actions that take place within the context of entities in specific temporal and spatial locales; and contingencies, which are actions in the context of the relationships between events.

Developed and refined over several years, the EEC framework is ideally suited for this analysis. It has a natural connection to the mediation model of associations, as contingency graphs can be translated into associograms. Practical application of the EEC framework is still ongoing, and much of the software development has run in parallel with my research.

Described in greater detail in (Suthers & Rosen, 2011), the EEC framework was built in part to help study activities within sociotechnical systems. The EEC system was designed to be flexible enough to incorporate any number of events that can take place, the relationships between events, and the entities involved. The toolkit used to implement the EEC was developed in Java, utilizing a database layer managed by Hibernate[3], which supports the storage and retrieval of Java domain objects. In short, server logs are imported by scripted processes and stored as EEC objects. Entities and events, as defined by the user, are stored as base EEC objects. Contingencies can have a pre-defined description and be created during the import process or added on post-import once entities and events have been stored. Abstraction techniques can be used to identify and create different aggregated groupings, constituting *composites*. Analyses can be conducted on any dataset by defining parameters with which to view the data. For this research, association-specific connections can be made between entities via different events. Application of the EEC tools are currently being used for small group chat analyses (Suthers & Desiato, 2012).

[3] http://www.hibernate.org/

4. Methodology

My experience from the previous studies has helped provide a better understanding of the relationship between actors and technology in a sociotechnical system. Our work has contributed to studying boundary spanning, and how different media artifacts can play an important role in member activity. Furthermore, I have developed a sound concept of the methods needed by using ESDA techniques to find abstractions from log data that can assist in studying sociotechnical environments.

4.1. Managing raw data

The first step in the analysis began with the raw log data from a sociotechnical system. This data can be generated by user activity within a website. Whenever a user submits information in a textbox, clicks on a hyperlink, opens or closes an area with text, selects form elements, or performs any other number of actions, the server logs will have recorded the act. Each log entry is also embedded with other data that provides additional context and information about the action. In general, besides the URL that directly corresponds to an action, server logs can also include timestamps, session identifiers, referrer URL's, virtual locations, and other relevant information.

Unfortunately, raw log data is not very usable for most types of analyses. Most of the data is unintelligible, e.g. timestamps are widely kept in Unix time format, or number of seconds since January 1, 1970. Session identifiers are randomly generated strings of numbers and letters. Even a straightforward URL can complicate matters because they do not always have a one-to-one relationship with user actions, e.g. a single user action such as a mouse click can potentially generate multiple URL's to be logged.

4.2. Chunking and coding events

The next step is to apply several different smoothing operations to organize the raw data and make it usable.

Chunks *are aggregates of adjacent data elements that the analyst views as coherent. Chunking tends to be used most heavily at the outset, when analysts are faced with an unrestricted mass of events whose only connection with one another is their temporal relation.* (Sanderson & Fisher, 1994)

A chunking operation is the logical process to apply to raw log data. This helps solve the many-to-one relationship between log entries and user action events. Given a specific key of what URLs belong together, fast filtering passes can be performed which will properly group the URLs that correspond to a single user action.

Once the logs have been properly chunked, all of the entries can be coded.

Codes *are syntactically structured labels that are usually linked to data elements or chunks* (Sanderson & Fisher, 1994)

A URL can be difficult to interpret, and often no more legible than the random string of text in a session identifier. Codes may be predefined before the actual operation, but will vary with different applications. I codify each log entry into human readable descriptions that any analyst familiar with the system can understand. This operation is also necessary immediately after chunking because once several log entries have been chunked together, the aggregated set no longer has a unique identifier; coding the newly created group provides a formal label so that it can be referenced without confusion.

Through two ESDA processes, chunking and coding, I have created the necessary abstractions from the original log files. The new data has resolved the problems of server log ambiguity and illegibility. It now provides a chronological list of events that are understandable. For example, the following is an abbreviated sample of entries in a disCourse log file:

Table 4.1: Example disCourse log sequence

	URLs
1	/workspace/418/note/2980/discussions/tree/1096
2	message-toggle-raquo:16161
3	reply-form-opened:16161
4	/workspace/418/note/2980/messages/reply_preview/16161
5	post-reply:16170

Most column data (e.g. timestamp, user ID) have been removed except for the URL request. After chunking and coding the entries, the resulting data now looks like:

Table 4.2: Events created from log data

	Events
1	View discussion 1096
2	View message 16161
3	Reply to message 16161 with message 16170

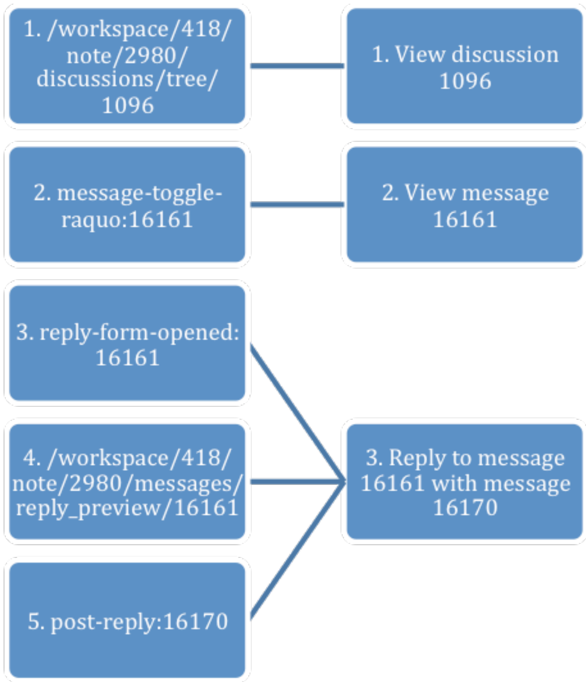

Figure 4.1: Chunking log data (left) into events (right)

The transformation is also illustrated in Figure 4.1. URLs 1-2 from the raw data have a one-to-one relationship with their codified counterparts in the events (events 1-2). A careful comparison of the translations would reveal exactly how the parsing and filtering processes worked. Lines 3-5 from the URLs column were chunked and coded into line 3 of the events column. It is more difficult to see this translation. A user clicked a message reply button (log #3), entered text in a form (log #4), and clicked another button to submit the form (log #5). All of those actions resulted in a reply event, as shown by line 3 in the events column.

4.3. Abstracting events into interactions

After creating a set of events, I can begin to combine them to form additional higher-level abstractions.

Connections *represent either the relations between ESDA products of a similar type of links between qualitatively different ESDA products that are nonetheless based on the same data elements. First, because of the situated, event-driven, and often haphazard nature of real-world events, topics, intentions, and themes do not necessarily emerge in a linear, temporally contiguous way but instead emerge in nonadjacent parts of a record.* (Sanderson & Fisher, 1994)

Unlike the chunking operation, events being connected do not have to be adjacent to each other, and in many cases are not. Additionally, events are not merged into a higher-level description; they each maintain their individuality. I connect events to find *interactions*. I am defining an interaction as a sequence of at least two separate events, adjacent or non-adjacent, between two different users that involve the same artifact. For example, if a user creates a wiki page, and another user reads the wiki, I say that user B (the reader) has interacted with user A (the creator), as mediated by the wiki (the artifact). Furthermore, each interaction has additional properties that help define it. In the wiki example, the interaction between user A and user B is recent (temporally close) and unidirectional.

The description of the interaction is important because the interactions between users form an important unit of analysis in these studies. There are many different

Figure 4.2: Events that can build uptake

forms of interactions, and how the various technological artifacts mediate them in a system is the focus of this research. By tracing the interactions between two users, I can connect multiple events – which are in chronological order but not necessarily adjacent – and identify those that are contingent upon earlier events. Contingencies are observed (empirical) relationships between events that indicate how one event (action) is contingent upon prior events. The concept of contingencies and how they used to identify interactional sequences is derived from (Suthers, Dwyer, Medina, & Vatrapu, 2007; Suthers et al., 2010). To elaborate further on the previous wiki example, user B can choose to edit the wiki with his own contributions after reading it. There are now three events that can be connected as in Figure 4.2.

This sequence is an example of *uptake* (Suthers et al., 2010). Suthers describes uptake as "when a participant takes aspects of prior events as having relevance for ongoing activity." In Figure 4.2, events 2 and 3 are contingent upon event 1 having already occurred. The sequence of events between these two users suggests that user B has taken some idea from what user A posted, and then contributed additional knowledge to the original artifact. The contingencies between the events can be evaluated for uptake. An argument can be made that the content of user B's wiki edit was influenced by his reading of user A's original wiki post. Clearly, a more conclusive statement cannot be made without additional information (e.g. interviewing user B, or analyzing the content). However, using only log files, I can still assert that the data supports the occurrence of uptake in this scenario. This serves as a simple example of how the ESDA operations and abstraction of events can support a higher level of analysis.

4.4. Creating associations from interactions

Having a grasp of interactions between actors and contingencies between events, I can now step ahead to build the mediation model of the analysis, an associogram. Associograms can be constructed from individual events without any relationship. As events in this framework includes information on two actors mediated by an artifact, it provides the fundamental components of an associogram tie. For example, given an event where user B reads a wiki page created by user A, we have the ingredients necessary for a visual representation of an association, i.e. the *actor-artifact-actor* triplet. This is a straightforward process, and allows for easy creation of associograms with minimum data.

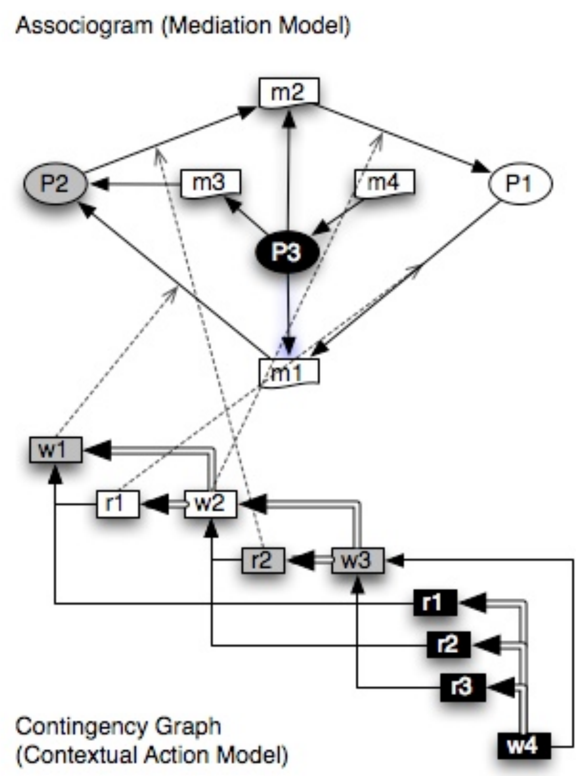

Figure 4.3: The relationship between associograms and contingency graphs. From (Suthers & Rosen, 2011)

Associograms can also be constructed using from a contingency graph. I present an example using Figure 4.3. The bottom portion of the figure represents a contingnecy graph. Each rectangular node is an event, and arrows denote contingencies that are directed from a contingent event to the source event. The various shades represent three different actors, in white, grey, and black. "W" is a write event, and "R" is a read event. We convert this contingency graph into the associogram in the top portion of the figure. Each event node is compressed into an associogram tie; the tie can be made up of many different events. The nodes in the associogram represent actors ("P" is a person), and correspond to the three colors used in the contingency graph. The mediating artifact, hidden in the contingency graph as a component of an event, is made explicit in the associogram. The dotted arrows are examples of the relationships between the two model views. This process is more complex than using pure events, but offers the ability to create easier automated conversion processes. Unfortunately, the associogram does not clearly account for temporal components; this information, "stored" in each contingency arrow, can be lost in an associogram representation (e.g. in Figure 4.3, we no longer know if P2 posted m2 first, or if P1 posted m1 first).

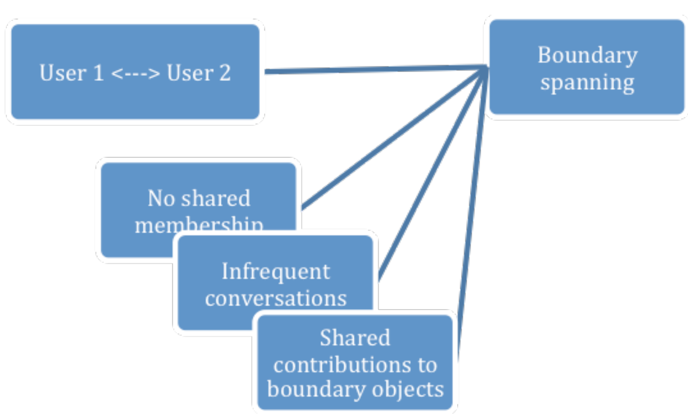

Figure 4.4: Defining a tie by different patterns

4.5. Computing patterns

While a tie can be constructed purely out of associations between two members, it is significantly more valuable if we can identify how it relates to other patterns of interaction.

Computations that expose particular patterns signify certain types of ties between members. From ESDA, “Computations refer to the broad range of formal procedures for analyzing sequential data that different ESDA techniques offer.” (Sanderson & Fisher, 1994). I begin the computation process by obtaining the different measurements based on pair-wise association data. These properties will help define ties by more than strength (strong or weak), but with richer information defined by various patterns (e.g. uptake) from which they are constructed. For example, using two random members, rather than claiming user A has a strong tie with user B, we can use their association data to describe an infrequent, synchronous chat based relationship via boundary objects from different groups, but not having any shared membership. This might offer hints to a potential boundary spanning relationship.

4.6. Summary

The proposed study is an exploratory analysis of user interactions within a sociotechnical system. The method of analysis is founded on the results of earlier pilot studies, and applying many of the techniques following ESDA. By applying these principles, I aim to create a high-level meta-view of the actors and their associations in the system. This offers an opportunity to establish the connection between the various technological media to the different associations that are formed.

I have reviewed literature that covers different aspects of social network analysis, including evaluations of dynamic social network analysis, the strength of ties, and boundary spanning. Because my research questions are focused on how online interactions are embedded in technology, I also focused my attention on sociotechnical capital. I have reviewed different methods of research, actor-network theory and exploratory sequential data analysis, all of which provided insight on developing the methodology and framework to be used. The two pilot studies have established a connection between technology affordances and activities such as boundary spanning, and offer hints of possible paths to further develop the research. The EEC framework offers an opportunity to study online sociotechnical systems in a manner that can help understand the complex relationship between technology, the types of associations formed between actors, affordances for interactions, and the sociotechnical capital connected with those interactions.

4.7. Data Source

In this section, I describe the community that I will be using for my study, the Tapped In network, and how I prepared the data for analysis.

4.7.1. Tapped In

SRI International’s Tapped In is an international network of educational professionals that are engaged in diverse forms of informal and formal professional development and peer support (Farooq, Schank, Harris, Fusco, & Schlager, 2007; Schlager, Fusco, & Schank, 2002). Cumulatively, Tapped In has hosted the content and activities of more than 150,000 education professionals in thousands of user-created spaces that contain threaded discussions, shared files and URLs, text chats, an event calendar, and other tools to support collaborative work. Through Tapped In, organizations can develop, implement, and manage online courses, workshops, seminars, mentoring programs, and other collaborative activities that supplement or function in lieu of face-to-face activities. More than 50 "tenant" organizations, including education agencies and institutions of higher education, have used Tapped In to meet the needs to students and faculty with online courses and workshops. Approximately 40-60 community-wide activities per month are explicitly designed by Tapped In volunteer members to help connect members, with groups forming organically after members meet in these activities. Extensive data collection capabilities underlying the system have captured the activity of all members and groups. The Tapped In codebase is available under an open-source (GPL) license.

SRI colleagues have provided eight years of anonymized data for our studies. Out of this data, we selected a period of peak usage that occurred from September 2005 through May 2007 for analysis.

4.7.2. Data Preparation

The raw data from Tapped In required several iterations of data manipulation in order to prepare it for use.

4.7.2.1. Database information

Several database tables were used for the study. Each table is described in detail below, including the relevant columns of data that is used in the study (the actual column name will be in parenthesis and follow the description, case sensitivity is based on original naming schemes). These columns are not comprehensive, as other data also exists for the tables but is not used.

disc_message: Each message that was posted in a discussion forum is stored here. Each row of data includes the message identifier (MSG_ID), a parent message if it exists (MSG_PARENT_ID), the discussion that the message belongs to (MSG_DISCUSSION_ID), the person who created the message (MSG_OWNER_ID), and the timestamp when the message was created (MSG_CREATED).

discussion: Every discussion in Tapped In is stored here. Relevant information for each discussion includes an identifier (DISC_ID), the location where the discussion exists (DISC_ROOM_ID), the user who owns the discussion (DISC_OWNER_ID), the root message in the discussion (DISC_ROOT_MSG), and a timestamp of when the discussion was created (DISC_CREATED).

room: Information for all the virtual rooms in Tapped In is stored here. Rooms are virtual spaces where many of Tapped In's activities take place. Room data includes an identifier (ROOM_ID), nickname (ROOM_NICKNAME), an owner (ROOM_OWNER_ID), and type (ROOM_TYPE).

tappedin_user: All Tapped In users are stored in this table. The personal data has been anonymized for privacy. The usable data includes a numeric identifier

(ExpUid), the user's virtual campus (HomeCampus), and the primary organization they belong to (PrimaryTenant).

uploadfile: All files uploaded to Tapped In are stored here. File information includes an identifier (FILE_ID), the room the file resides in (FILE_ROOM_ID), the creator (FILE_CREATOR_ID), the mimetype (FILE_MIMETYPE), and the creation time (FILE_CREATED).

web_request_hist: All user actions are stored here, including every user mouse click. Each row contains the username (username), location of where the action took place (room_nickname), the type of artifact being interacted with (item_type), the specific artifact involved (item_id), and a timestamp (click_ts).

Figure 4.5 is a simple visualization of the table relationships.

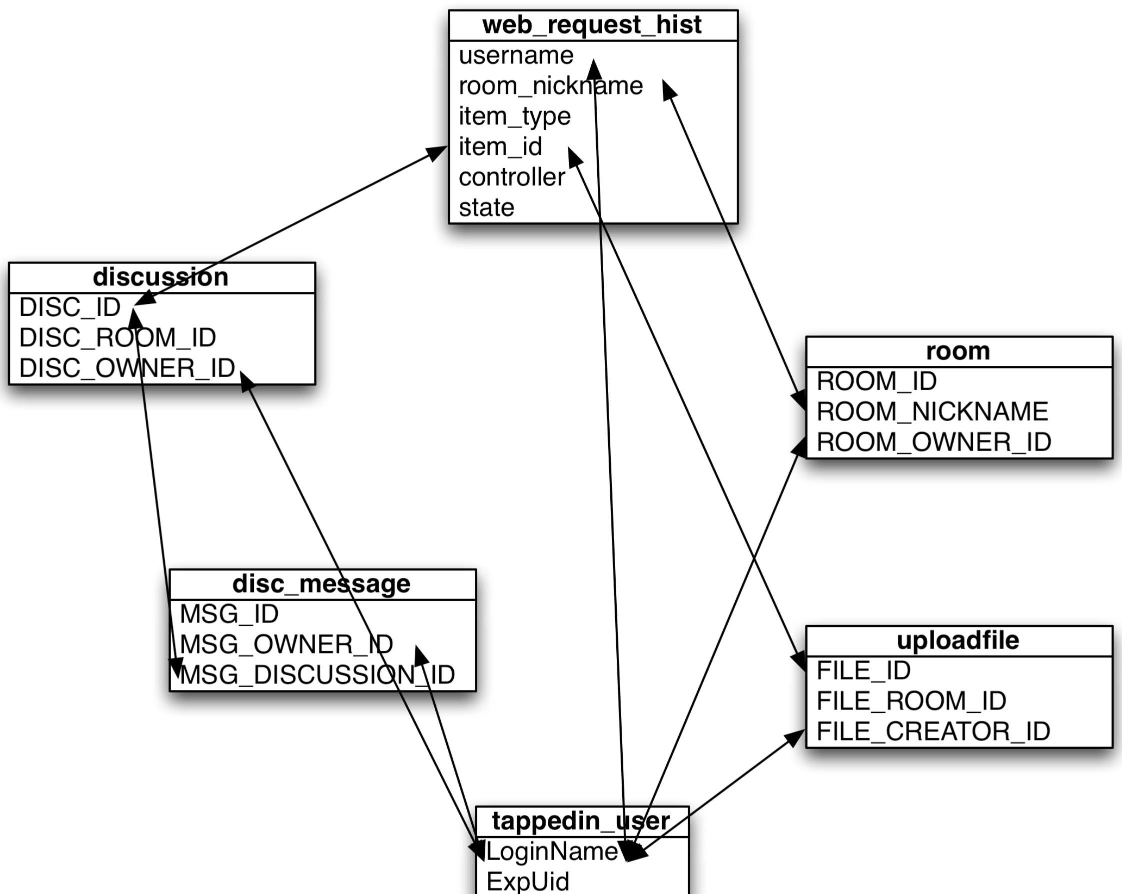

Figure 4.5: Relationships between Tapped In database tables

Data from the above listed tables were adjusted to make them usable for the study. Most of the scrubbing processes were motivated by incomplete information. Some of these processes are now described.

4.7.2.2. Preparing Events

The *web_request_hist* table contains all user actions and URL requests by the system. The first issue in the web_request_hist table was portions that had incomplete information. For example, in each case where the action was to create a file, the information stored includes the user, artifact type (file), location, and timestamp. However, the identifier for the file (item_id) is not included in the log. It is likely that the log stored the request to create the file before the actual creation event takes place. To resolve this, the username and timestamp of every file create action was used to match up against all records in the uploadfile table. Because the uploadfile table contains the creation time for each file, I was able to matchup the create record with the associated file. An additional issue arose where the timestamps did not match the exact time recorded to the second, so a buffer of plus/minus three seconds was used to match the records.

4.7.2.3. Preparing Users

The user list from the most current database snapshot contained users that were still in the Tapped In system at the time. However, it is not surprising that many users have been deleted or became inactive during the history of the system. Users who have been inactive for over a year are also regularly cleaned up and removed from the system; thus, there were many activities during the two-year period of study which were conducted by users who were no longer in the system. I was able to reconstruct the user table by incorporating data from deleted users, and several snapshots of the database as it existed in the past.

Two major user groups were excluded from this study. Students from a K-12 virtual campus, as activities from K-12 students were not covered by the Institutional Review Board for this research. Guest accounts were also removed, as they can be reused by different people, and there is no method of tracking a unique user. Additionally, their activity tends to be very short, which does not provide much usable data. Both student and guest accounts were deleted from the user table before any analyses.

4.7.2.4. Preparing Chat logs

Logs from live synchronous chats were stored as text files, grouped by the virtual rooms where they occurred. This immediately caused several problems in how to incorporate the data into our framework.

1. The chat data must be imported separately from all other data, which existed inside the same databases.
2. The plain text files had no means of separating the types of data stored. Each text file contained chat logs that recorded a timestamp, username, and contribution.
3. Timestamps were a major concern. Since logs were stored as plain text, there was no hint as to the context of time zones, or any adjustment to daylight savings, and logs used time stamps from different time zones.

4. Usernames caused problems because Tapped In's system allowed a username to be reused if the original had been deleted. Inside the database tables, only a numeric user identifier was unique.
5. Contributions came in different formats. There are events when members join or exit, emoticons, actions, or plain "spoken" text.

All of the problems had to be resolved separately before the data could be imported and joined with other database tables. Each resolution is as follows:

1. A new table was created to store all chat data, separate from events recorded in other tables.
2. A parser was created to separate the different data types into separate fields. Regular expressions were used to identify the timestamp, usernames, and contributions.
3. Timestamps were compared and matched up against known events stored in other database tables. For example, the table *chat_session_hist* was used specifically for this purpose, as it included timestamps of when members would join or exit chat rooms (although it did not record contributions).
4. Each username logged in the chat files were compared against any username that matched, regardless if it had been deleted or not. Usernames were only considered to be legitimate if all chat actions occurred between when the user account was created, and before the date of the user's last activity. Additionally, the user identifiers were added to the final chat data table once the username was verified.
5. A parser was created to apply regular expressions that identified whether a contribution was a join/exit action, a user specific action, or text dialogue.

4.8. Addressing the research questions

The focus of this research is to explore how technology mediates relationships formed in sociotechnical systems. I have outlined my proposed method of analysis: obtain Tapped In event data, apply techniques using the EEC framework, create abstracted associations by aggregating interactions, and then identifying patterns to determine the mediating effects on actor relationships. These activities relate to the research questions as follows:

- On interactions within the sociotechnical system: These can be identified by actors' read and write actions on the different digital artifacts available in Tapped In. We can also distinguish between artifacts based on their different affordances.
- On relationships: These are expressed by the pattern of interactions in actor associations. I can classify these patterns, and potentially identify groups of similar patterns to see the similarities between how actors interact.

- On technology's enabling effects: Having connected digital media with classified groups of associations, I can determine the role, if any, that technology plays in enabling different types of relationships.

Of the three sub-questions, the last one is the most open ended. The first two questions can be answered by studying the Tapped In system. After importing log data into the EEC framework, creating the necessary abstractions into associations, and identifying different patterns of interactions, we can describe both the sociotechnical interactions and actor relationships. However, to answer the third question, and following that, the main research agenda, I need to focus deeper into connecting the effects that digital media, and its affordances, are linked to the relationships. This will help understand whether, and how, the relationships between actors are mediated and reflected by the technology they use to interact.

4.9. Human Studies

The study of human subjects in this study was determined to be exempt from Department of Health and Human Services regulations. This exemption was granted and can be referred to under Assurance Identification number F-3526 and Institutional Review Board (IRB) registration number IORG0000169. The certificate of exemption can be reviewed upon request.

5. Multi-modal Multi-granular Analysis

5.1. Introduction

This section describes work that is a preliminary study applying some of the methods and framework discussed earlier. It is not the main topic of this research, but I introduce it prior to the primary study to provide some context as to how we can examine artifact-mediated associations. It presents a multi-modal multi-granular (MMMG) method of analyzing network and interaction data, and how different levels of granularity reveal unique properties of the network. We utilized various techniques and visualization tools to bridge between levels of analysis and study multiple digital media under different lenses. The analytic techniques are explicated via analysis of interactions in SRI's Tapped In community of educational professionals.

As each sociotechnical system offers unique tools to support interactions, we need methods that can describe how users are connected to each other via the many different ways they interact. We also want to account for the different granularity with which we view data regarding these interactions. From a high-level "forest view" of communities as a whole, down to individual actions that help foster relationships between users, we show how the multiple layers of granularity are distinct in presenting different frames of reference, but also connected in how the different perspectives can be related.

The summary of this study is based on (Chu, Rosen, & Suthers, 2012).

5.2. Analysis

From a previous study (Chu, Suthers, & Rosen, 2012), we selected a random sample of events that showed how Tapped In members were using the system. The full dataset was not displayed because the amount of nodes and edges were too large to be handled by the SNA graphing software that was used (we alternated between UCINet and Pajek at various times), so our final results were randomly trimmed down to a viewable size. We constructed a macro-level associogram to identify which artifact types members were generally interacting with. We focused on four particular types: files, discussions, calendar events (the asynchronous meta-information related to a live synchronous chat session), and notes. Figure 5.1 visualizes the distribution of how Tapped In users were using the system to interact with each other. The results showed that the majority of Tapped In users focused on the file

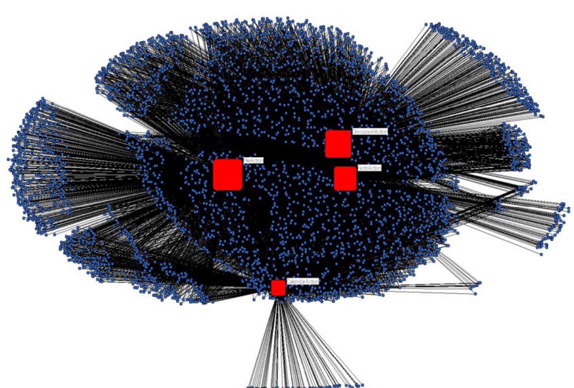

Figure 5.1: Associations between actors (small blue nodes) and four artifact types (large red nodes). Node size is degree

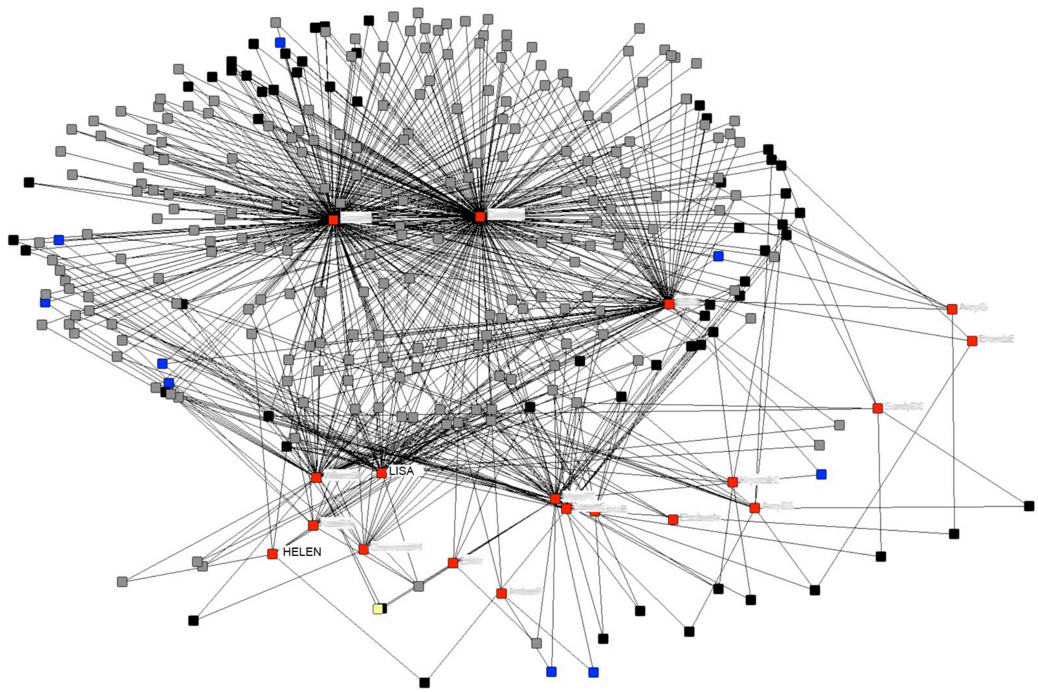

Figure 5.2: Multi-modal associogram of artifact instances, prior to focused chat session. Users are red, and various artifact types are represented by other colors. Discussions = blue, events = black, files = grey, messages = pink, resources = green, job links = yellow

sharing, wikis, and discussion forum components of the system (the larger three red nodes), and that calendar events were mainly created and moderated by a smaller group of members (the smaller red node toward the bottom). While this associogram is not the primary subject of this current study, a more detailed analysis can be found in (Chu, Suthers et al., 2012).

Our initial exploration was focused on finding "interesting" chat sessions within Tapped In. Several metrics were investigated to determine how interesting a session was, including number of members present, number of messages posted, etc. One particular chat session[4] was selected for further study because it exemplified the kind of engaged professional discourse the Tapped In environment was intended to support. Various analyses were conducted on this chat session. First, by examining several artifact instance level associograms from different periods of time, we can gauge whether the chat session participants varied in their mode of interaction over different time frames. If any such changes exist, we can begin increasing the granularity of our analysis to try and uncover the specific pattern of interactions that might have led to the change in behavior, and identify the artifacts involved with these interactions. Knowing both the mediating artifacts and how the members interacted with them will help in defining the type of association that the members have, which can better serve us when applying other SNA methods of analysis.

Figure 5.2 and Figure 5.3 show artifact instance associograms between members that have participated in our selected chat session, called "Teaching teachers."

[4] http://tappedin.org/archive/transcripts/profdev/2006/20060406teachingteachers.pdf

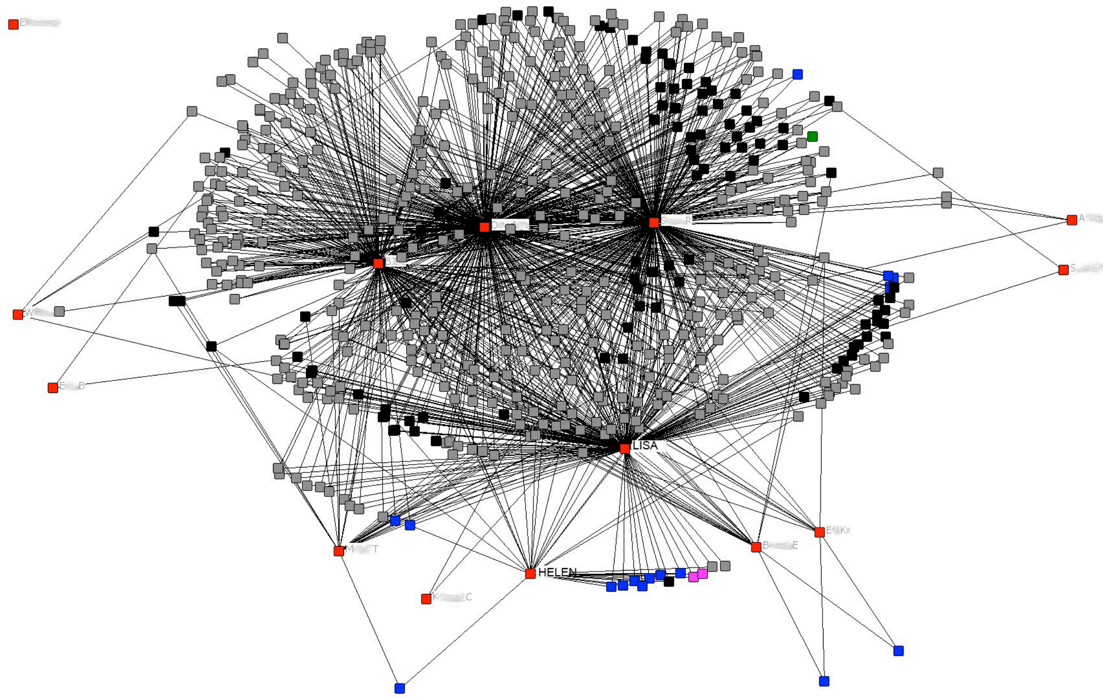

Figure 5.3: Associogram between users after focused chat session

Figure 5.2 shows all interactions that members have had with each other prior to the event; Figure 5.3 contains all of the interactions that occurred after. The two associograms provide a limited network level view of how the mediating artifacts affect interactions. After examining the two associograms, we chose two members whose interactions appeared to change after our target event, Lisa and Helen (these are pseudonyms). Importantly, Lisa and Helen began using discussion forums to communicate after the chat session had taken place. The first step was to identify all artifacts that had mediated any interactions between Lisa and Helen over the period of study. For example, files that were uploaded by Lisa and downloaded by Helen, or threaded discussions that both had viewed, chat sessions that both participated in, etc. The type of association with each artifact was not taken into consideration at this point; that is, a file uploaded by Lisa and downloaded by Helen was included, but so were files created by a third party that both members downloaded.

The event data for the two years of interaction between Lisa and Helen were then separated by month. We found that there was a spike in activity during July 2006, with higher levels of forum discussion and chat session activity than other months, and so focused our study on those events. From this point, we increased the resolution of our analytic view, now exploring individual events performed by each member. We constructed a contingency graph (Figure 5.4) with all of the events to visualize the different contingencies between all the events during the specific period of interactions.

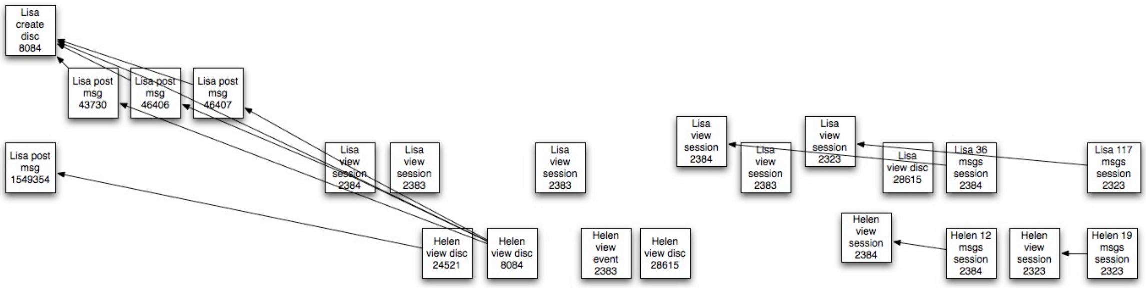

Figure 5.4: Contingency graph of Lisa and Helen interactions in a 1 month period

5.3. Results and Discussion

The association between Lisa and Helen can be summarized by their shared interactions. Both users have participated in two shared chat sessions. During the first session, Lisa contributed 36 messages while Helen contributed 12. In the second session, Lisa contributed 117 messages and Helen contributed 19. They have also shared interactions in two discussion threads. In one, Lisa posted three messages that Helen read. In the second, Lisa posted one message that Helen read. In both cases, Helen did not post any message. These mediated interactions suggest that the two members likely have a producer-consumer relationship where Lisa generally contributes content that Helen accesses. We can visualize this association in Figure 5.5. The two circles represent the two modes of interaction between Lisa and Helen. The thicker bottom arrows represent the mediated association via chat sessions. The direction of the arrow is analogous to state dependency, where the state of the discussion artifacts is dependent on Lisa's contributions, and the state of Helen is depend on having accessed the discussion artifacts. The thickness of the arrow represents the frequency with which each member has interacted with the artifact. The bottom arrows represent associations with chat session artifacts. The double arrow headers denote that each member's state is dependent on having accessed the chat session, and the chat session's state is also dependent on contributions made by the members.

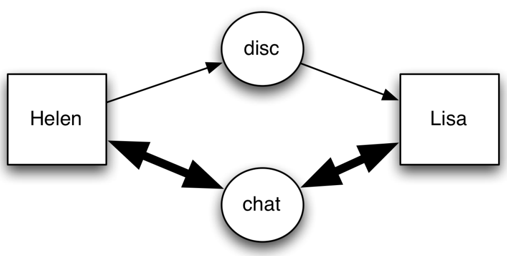

Figure 5.5: Associogram between Lisa and Helen as mediated by chat sessions and discussions

We can further compress the associogram into a sociogram by removing the mediating artifacts (Figure 5.6). This will reduce the amount of information provided, as we no longer know what type of artifact is mediating the tie between the two actors. However, we can still retain the different flows of information by varying the size of the arrowheads. Typically, the arrowhead in a sociogram tie represents a directional relationship, such as one person being friends with another (Wasserman & Faust, 1994). Because we are compressing association data into the sociogram, the arrowhead here represents the level of which information is flowing from one node to another. The direction is the opposite of the associogram dependency, as it now represents information flow. In Figure 5.6, information is flowing in both directions between Lisa and Helen, but it is greater coming from Lisa to Helen.

The multi-modal multi-granular approach has helped provide insight into several aspects of my research. Beginning with a single chat session (chosen after reading other professional development sessions), we began an MMMG analysis on the participants by comparing associograms before and after the chat session. By changing granularities, we discovered two members whose interactions had changed over time; a further look at their monthly activity levels helped narrow down time periods where other events of significance might be occurring. The patterns of interaction identified in the contingency graphs have revealed the type of relationship that Lisa and Helen have with each other, from within the context of the multiple media used to interact. The network tie between the two actors now contains a richer body of information, and will be useful for social network analysis when combined with data from other ties.

Figure 5.6: Compressing the associogram into a classic sociogram

The results in this study demonstrate the utility of a multi-modal multi-granular analysis. The association between Lisa and Helen is not defined by a single mode of interaction. The two members take advantage of several different digital artifact types at various times to communicate. Under close inspection, we found that Lisa and Helen had a consistent pattern of interaction, where Lisa was generally a content producer and Helen was the information consumer. Using the patterns to help understand the associations between members provides strong clues about the nature of their association. This is not to say that we have defined the two as a general producer or consumer; we are only examining interactions that pertain to a specific set of events that occur between them. Therefore, our conclusions are describing an aggregated form of a tie between those actors, and not a general role they play in the community as a whole.

The multi-granularity portion of the analysis has shown that viewing data under varying resolutions can reveal findings that are not easily visible. At the coarsest level, we are deprived of certain variables (e.g. time) but can identify structures in the network. These structures can help identify which artifacts attract the most or least activity, how centralized a network is, and possible sub-clusters that warrant additional investigation. Results from a previous study helped identify entry points showing how new members started using the Tapped In system (Chu, Suthers et al., 2012).

We have also demonstrated the importance of the meso-granularity view. While no longer capturing the network structure of associations, it reveals the specific instances of each artifact, and offers much higher resolution into identifying exactly what artifacts might be of interest and warrant further study. We can see individual files, discussion threads, or chat sessions in Tapped In have certain properties that might be mediating “interesting” interactions. With other systems, this might apply to member profiles, private messages, blog comments, or any other number of mediating artifacts. In many cases, the meso-level offers a transition to fine analyses such as contingency graphs, but it also provides its own unique information, such as understanding the association of small groups of individuals as mediated by certain artifacts. For example, we were able to define the specific pattern of interactions that connect two people, as we have seen with Lisa and Helen.

Micro-granularity is the finest level of analysis in our study. Here, we can look at both content and patterns of interactions as seen in the contingency graph. Tracing out a timeline of contingencies helps us identify the patterns of activity. Additional content analysis can shed light on how users might respond to each other differently under varying settings. It can also better describe user associations in small groups based on how they communicate. There is parallel research that is following along these paths (Suthers & Desiato, 2012).

I learned several important concepts in this study that informed how I next approach the data from Tapped In. The concept of multi-mediated associations is the most important. The different ways through way people interact can give us important clues as to the nature of their relationship. Both the type of media and the direction of interaction were necessary in helping to identify how Helen and Lisa relate to each other. This study only focused on a small window of time; extending the scope of analysis should give a broader picture of the relationship between individuals.

The research has shown that the changing the granularity of the study is a valuable method. It provides a connection between coarse and fine views of different representations of the data, helping us to see local interactions emerging into global phenomena, and how to find points in large-scale views to focus on. Additionally, the meso-level analyses with associograms provide perspectives on the data that would not be visible in large sociograms or single interactions. This understanding actually contributes to the emphasis on how to proceed with my analysis. The mediation model is an ideal approach in understanding the link between the social and the technical. Information regarding the mediating artifacts provided valuable information about Helen and Lisa, in both understanding their relationship and how they are interacting. Given they are a single dyadic pair, this study has shown that I can perform the meso-level analysis across *all* pairs of actors in the system to help answer the original research questions of this research.

6. Association Study of Tapped In

I conducted a second analysis using the same Tapped In data as in the previous study reported in Chapter 5. Based on that previous study, I focused on two specific components: collecting association data for dyadic pairs of actors, and identifying the differences in mediating artifacts. Having established a working method of creating associations between pairs of actors, I now automated the process to create associations across all possible pairs of Tapped In members. The associations, which aggregate information on each member's pattern of interactions, can then be grouped using statistical clustering algorithms. In identifying clusters of actor pairs with similar interaction patterns, I explore the actions being conducted by specific pairs and discover whether there are similarities in their behaviors. These results can answer the major research questions of this study – whether, and how, technology affects and mediates relationships formed in a sociotechnical system. In this section, I will describe the process of how the data was used to create associations, the clustering methods applied to group all the dyadic pairs, specific examples of the interactions between pairs in different groups, and the conclusions that can be made regarding the connection between technology and the relationships between members.

6.1. Methods: Data organization

The organization of Tapped In data was managed in a similar manner as the previous study. As before, event log data was imported into the EEC framework to establish all of the entities and events involved. Contingencies were not used at this stage, as they provide information at a finer granularity than would be necessary. The artifacts that I compare are files, discussions, and chats. Each artifact has two types of interactions – create and access – and has a different method by which the interactions are counted and recorded.

6.1.1. Artifacts

Files are any non-image files that have been uploaded to the Tapped In system. Images were filtered for two main reasons: first, because they are generally used only in personal profiles rather than supporting interactions between members. Second, the Tapped In logging mechanism considers any view of a member profile image as a file download action, even though the intention of the member is to view a profile and not download a file. For perspective, there are a approximately 221,000 file download events, of which about 116,000 (75%) involved images. While it would be impossible to check each image file to determine whether or not it was a profile picture, it nonetheless reflects the overwhelming nature of image downloads as compared to downloads that are definitely deliberate. Removing image files leaves 5853 files created during the study period, divided among 11 types. These are (including counts):

Table 6.1: Distribution of file types

Word document	4052
PDF	725
Powerpoint	325
Generic binary	314
Excel spreadsheet	168
HTML	121
RTF	82
Plain text	33
Zip	28
Quicktime	4
RealAudio	1

File interactions are simple and symmetrical. Creation interactions are counted when a member that has uploaded a particular file; access interactions are counted when a different member downloaded a file. In an associogram, the nodes are connected as seen in Figure 6.1. User X has downloaded File Y from User Z. When compressed into a classic sociogram, it is represented as a directed tie between User X (downloaded) and User Z (uploaded). This can be read as "User X is dependent on User Z as mediated by File Y."

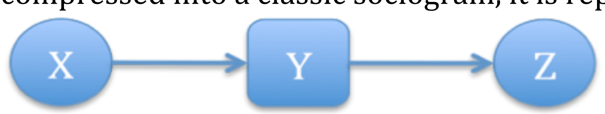

Figure 6.1: Generic associogram representation of different actions

Discussion artifacts are asynchronous threaded discussion forums. Users have the options to create new threads, create new messages within a thread, and then read messages other users have posted. Creation interactions are counted when a user has posted a new message in a discussion. Access interactions are slightly different. When a user opens a discussion to read, every message in that discussion is counted as an access event. The reason for this method is because of how discussions and messages are setup within Tapped In. Users will generally click on a link to a discussion to view the messages. The default system response is to open all the messages for viewing. This behavior eliminates the need for the user to click on any specific message. Therefore, the majority of user actions with regards to discussions are opening an entire discussion rather than individual messages. Based on raw event counts, the ratio of clicking on a discussion to clicking on a message is approximately 38-to-1. In the cases where users access an individual message, it is likely because they changed the default behavior and force messages to remain hidden unless explicitly clicked on. During our period of study, 12,136 discussion threads were created, containing a total of 75,976 individual messages. In an associogram, it looks identical to how a file association is comprised; Node X has read from Discussion Y that User Z has posted in (Figure 6.1). In most cases, each access event will be connected to multiple create events, as long as the discussion has more than one message. Compressed into a format for sociograms, User X has a directed tie toward User Z as mediated by Discussion Y.

The artifacts associated with **chats** are synchronous virtual rooms where many users can join. They can post and read messages to communicate with each other in real time. Chat sessions can be organized and prearranged, or occur spontaneously. Organized chat sessions are referred to as *calendar events*. Typical calendar events are usually an hour long, and take place in a predetermined virtual room. A user is designated as the facilitator of the event, who might or might not be the same person who created the event. In many cases, these events recur at specific time intervals. For example, the event "Tapped In Tips and Tricks" is a recurring event organized and run by Tapped In leaders, with the purpose of helping new users navigate and use the system. Chat rooms are also embedded into many aspects of Tapped In, e.g. all personal offices have a chat room. Consequently, impromptu conversations often take place without any planning.

Creation interactions are counted whenever a user posts a new message in the chat room. Access interactions are counted if a user is in a chat room when another user has posted a message. Unlike with files or discussions, this is a passive access, as no action is required by the user accessing the chat for the interaction to occur (other than entering the chat room). Because chats afford synchronous interactions, there is significantly more activity surrounding this artifact when compared to files and discussions, both of which are asynchronous. During our period of study, over 4 million contributions were made in chat rooms. In an associogram, the mediating artifact represents the chat room in which both users were present; Figure 6.1 represents the interaction where User X was present when User Z made a contribution in chat room Y. In most cases, each message posted will have multiple readers, unless there are only two people in the room. In the sociogram representation, User X has a directed tie toward User Z as mediated by chat room Y.

All user and room data were imported into the EEC framework. All events were also imported, along with artifacts including files, discussions and messages, and chat logs. This process created all the events and entities necessary to identify association information based on interactions.

Each event has several pieces of information contained in it. I am interested in two specific components, the actor and the artifact. Therefore, for every event that occurred, I derived an actor-artifact pair from the event. Following that, for each read event, I connected the artifact to the actor who originally created it, whether by uploading a file, posting a discussion message, or making a chat contribution. Now, I have expanded the pair to a triplet, reader-artifact-creator. These data were stored in three new tables, *file_associations*, *discussion_associations*, *chat_associations*, each containing information specific to an artifact.

6.1.2. Dyadic associations

To merge all data into a single usable table, artifact and directional information was compressed. A new table was created, called *dyadic_associations*. This table contained 8 columns of data:

- Actor 1: a unique numeric identifier
- Actor 2: a unique numeric identifier
- Files1to2: count of occurrences where actor 1 downloaded files created by actor 2
- Files2to1: count of occurrences where actor 2 downloaded files created by actor 1
- Discussions1to2: count of occurrences where actor 1 read messages in a thread that contained messages posted by actor 2
- Discussions2to1: count of occurrences where actor 2 read messages in a thread that contained messages posted by actor 1
- Chats1to2: count of occurrences where actor 1 was co-present in a virtual chat room as actor 2 made a text contribution
- Chats2to1: count of occurrences where actor 2 was co-present in a virturla chat room as actor 1 made a text contribution

Once the data had been completely imported into dyadic_associations, the table contained approximately 227,000 rows of data. Each row of data describes all files, discussion, and chat interactions between two members of the Tapped In network. In essence, each row is the final product of an association between two actors.

6.2. Methods: Cluster Analysis

The next step in the process was a statistical analysis of the types of associations that exist in the network. I am interested in whether members interact in a particular way with other member of the network. If I can identify the characteristics of how certain groups of members behave, this would greatly help in finding interesting or beneficial patterns of interaction that are occurring in Tapped In.

I used the statistical software package IBM SPSS[5], version 20, to conduct the statistical analyses. SPSS provides a variety of tools in helping to view data, and to apply many different procedures to identify relationships between variables and perform clustering analyses. Other techniques and models are available, although I focus mainly on variable relationships and clustering.

[5] http://www-01.ibm.com/software/analytics/spss/

6.2.1. Choice of Clustering Method

Generally, there are two popular methods by which clustering algorithms can be run, hierarchical and non-hierarchical. Each has inherent advantages and disadvantages. A hierarchical cluster analysis attempts to find clusters by comparing similarity measures between data points. The similarity unit of measure is a user choice and can be any number of possibilities, e.g. Euclidean distance. One approach is to compute every possible combination of distances between data points in an n x n matrix. Starting with n clusters, the two closest data clusters are paired together, now leaving n-1 clusters. Distance measures are re-calculated for the new cluster. This procedure is iterative, as new similarity measures are calculated and clusters continue to be combined to form new, bigger clusters. The algorithm is complete when there is only 1 cluster left. It is up to the user to determine when an optimal number of clusters have formed; usually, this can be accomplished by plotting the percentage of change after each iteration. This is called an *agglomerative approach*, as you are working "up" a tree, starting with n leaf nodes until you have reached a single node.

An alternative hierarchical clustering method is the *divisive approach*. Starting with a single cluster constituting all of the nodes, begin dividing the cluster until you have reached n clusters. Aside from choosing between agglomerative or divisive, a choice must be made to determine how the similarity or dissimilarity measures are taken. Various distance measures exist, and are critical in how the clusters will emerge. Additionally, how new cluster similarity measures are calculated is another user decision (e.g. distance between closest members, distance between average of members, etc).

The k-means algorithm (Lloyd, 1982) is a non-hierarchical method of cluster analysis. A major difference is that the researcher must decide how many clusters to create, the k value, before beginning the analysis. Generally, the k-means algorithm will begin with k possible cluster centers, which can be randomly selected. From those initial centroids, all the remaining data points will be placed into an existing cluster, depending on their similarity to the centroids. When all data points have been assigned to a cluster, a new centroid is calculated, which can be accomplished with different techniques. The data points are now re-assigned to clusters based on the new centroids. This process is iteratively performed until a predetermined end time, or if there are no more changes in cluster membership. Similar to hierarchical algorithms, there are user decisions (e.g. how to calculate centroids) that can change how clusters are found.

There are different advantages and disadvantages to both the hierarchical and non-hierarchical approaches. One matter to consider for hierarchical methods is the size of the data. Similarity measures are calculated from every data point to every other data point, creating a large n x n matrix. This approach has time complexity of $O(n^2)$, and does not scale well. Another concern is that when data points are joined to a cluster, it is a permanent assignment. Regarding the k-means approach, given a normal distribution of data, the algorithm can run in polynomial time. However, the requirement to predetermine precisely how many clusters to look for reduces its benefits; choosing a "bad" k can produce misleading results.

Unfortunately, the data from Tapped In exposes flaws in both the hierarchical and non-hierarchical methods of clustering. The dataset is extremely large, and I do not know the unknown number of clusters to be found. However, SPSS has a third method of cluster analysis called Two-Step Cluster Analysis. As its name implies, the algorithm operates in two steps. First, it uses only a single pass over the data (thus supporting large datasets) to form a set of pre-clusters that provides a reasonable guess as to the number of clusters that exist. This process helps reduce the "initial" number of clusters in order to create a smaller similarity matrix. The second step is creating the matrix and applying standard hierarchical clustering techniques to that matrix. While the final solution might not always be optimal, Two-Step Cluster Analysis offers solutions that allow for good descriptions, and the option to continue refining the analysis with its results.

6.2.2. Concerns and Resolutions

Several problems manifested during the clustering process. The first is data ordering for each pair. Up until this point, all of the analyses have been conducted under a network lens; that is, directionality can be seen. The notation used for a network description of ties is non-linear, so there is no problem created by directionality. In addition, network graphs, including associograms, the visual order of any two nodes is irrelevant since they can be moved without any consequence. However, when attempting to apply statistical clustering analyses, where every row of data contains the information for both actors, the order of the actors is vital to the analysis. The linear nature of a vector requires an arbitrary direction to be chosen. For example, suppose we are given a random pair of actors where their association data is completely one-directional, such that actor 1 (in the left hand column) contributes all file, discussion, and chat data, and actor 2 (in the right hand column) contributes nothing. Suppose there is a second pair of actors where the association data is also one-directional, such that actor 3 (in the left hand column) accesses all file, discussion, and chat data, and actor 4 (in the right hand column) contributes everything. Their data can be seen in Table 6.2.

Table 6.2: Example data

Actor	Actor	File1to2	File2to1	Disc1to2	Disc2to1	Chat1to2	Chat2to1
1	2	10	0	10	0	10	0
3	4	0	10	0	10	0	10

As the table shows, the two pairs have opposite values for their interactions. Intuitively, we can see that both actor associations are identical; both pairs have producer/consumer relationships, where one actor contributes all of the data that is consumed by the other. However, because the order of the pair was arbitrary, the data is ordered in such a way that a statistical analysis would identify the two pairs as complete opposites.

To resolve this concern, the data needed to be reordered before being imported into SPSS. The reordering process used a single artifact type to be used as the baseline. That is, for one artifact interaction pair (1to2 and 2to1), all actors would be ordered so that the greater number would always be to one side, specifically in 1to2. I chose to use chats for the baseline artifact because of its higher volume of interactions than the other two artifacts. However, this does not imply I place any higher importance in chat interactions, only that its higher raw numbers would be a stronger contributor to later clustering. Therefore, a check was made for every row: if Chat1to2 was greater than Chat2to1, nothing was changed. However, if Chat2to1 was greater than Chat1to2, all artifact column pairs would be switched, along with the actors. This process would maintain the integrity of the data ordering, but arrange the data in such a way that one column would serve as the baseline artifact interaction with which to compare other interactions. I later re-ran the re-ordering process using discussions as the baseline artifact. The clusters found were very similar to those found using chats as the baseline. This provided some validation that using a single artifact to provide context for ordering worked, and that choosing any given artifact would not badly upset the clustering results.

A second problem is unbalanced data. Immediately after the data import, I examined some basic statistics about the association data. Table 6.3 is the descriptive summary.

Table 6.3: Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Files1to2	227124	0	142	.15	1.944
Files2to1	227124	0	144	.03	.616
Discussions1to2	227124	0	4410	6.76	46.220
Discussions2to1	227124	0	4924	5.36	40.378
Chats1to2	227124	0	43934	36.14	258.484
Chats2to1	227124	0	22648	17.67	154.291
Valid N (listwise)	227124				

The differences between artifact interactions are particularly glaring. Where the maximum file interactions in either direction is only 144, the discussion maximum is 4924, and the chat maximum is an even higher 43,934. While this makes clear the low volume of file interactions, the small amount will easily be lost when compared against much greater values of discussion and chat interactions.

In this case however, SPSS offers a simple solution. During clustering analyses, we can apply the methods to the standardized scores, or z-score, of each interaction count. SPSS automatically calculates the z-score by subtracting the mean from the raw score, and dividing the difference by the standard deviation. Use of z-scores serves to eliminate the “advantage” that chats and discussions have over files, by considering the deviations from the mean for every interaction type.

A third problem from the data was the existence of a large group of low interaction user pairs. This was not unexpected; given n members, there are $\frac{n(n-1)}{2}$ possible combinations of pair-wise associations. Based on event logs within Tapped In, there were over 23,000 unique usernames that performed some action during the two-year period in the study, leading to over 260 million potential pairs. In practice, only a very small percentage are realized, but even within that set, there will still be many that were only formed by chance encounters. For example, when Tapped In members login, the default behavior places them in a “reception” area, which contains a public chat room. A separate study using the same data found that the reception room artifact had 18,810 edges extended to different users (Suthers & Chu, 2012). Many of these are likely no different than a real-world building reception, where we pass by strangers on our way to our respective destinations to different floors and offices.

A large low interaction group always dominated each clustering pass of the Tapped In data, no matter the algorithm chosen. The volume of this group continued to prevent any other clusters from revealing themselves, as it tended to “swallow up” smaller clusters. One possible solution was to remove all interaction data that does not meet a minimum activity level criterion, although that risks losing valuable data, and any baseline values would be subjective. I chose a second solution: to conduct multiple hierarchical passes manually, and allow the smaller clusters to naturally emerge from the low interaction group with each pass, providing more accurate results. After each pass, a subsequent hierarchal analysis was done on each of the resulting clusters separately from the others to reveal its internal structure. I stopped this manual decomposition after seeing sample data that was close to cluster centroids, seeing more unique differences between clusters, and a decrease in variable discrimination.

6.3. Results

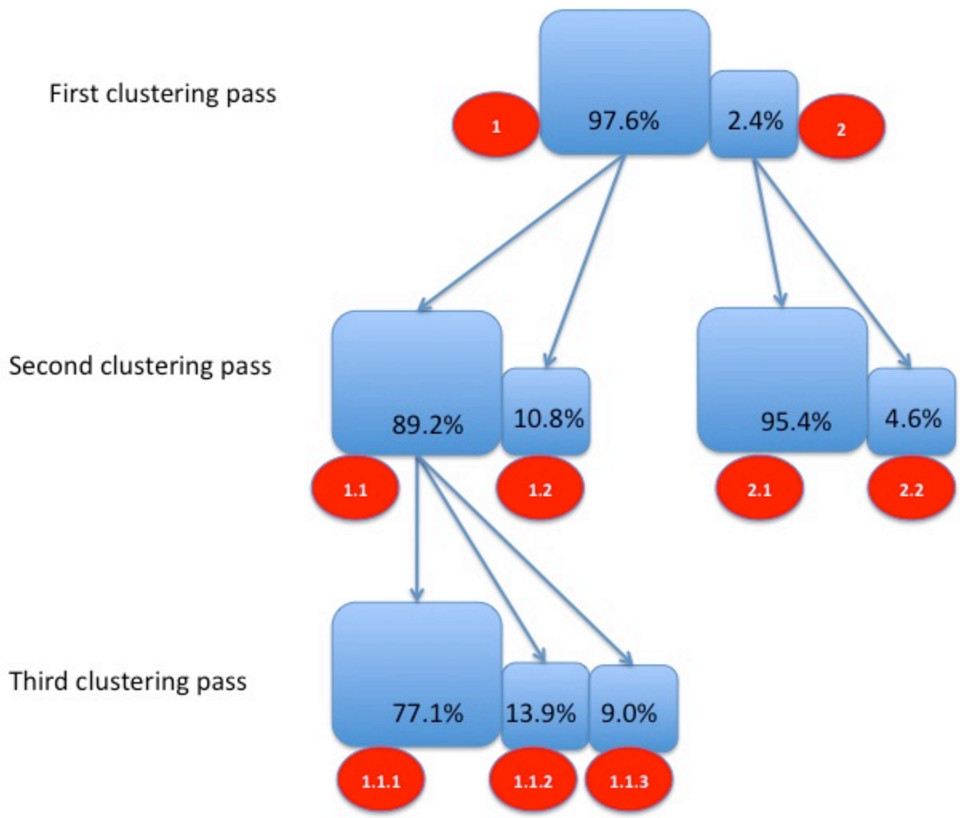

Figure 6.2: Tree breakdown of clusters, using Gorn numbering (Gorn, 1967). Blue nodes are clusters found, percentages are local for each pass, and red circles are Gorn name.

I completed three manual hierarchical levels of clustering; Appendix A contains cluster descriptions of the three passes, and Figure 6.2 is a quick summary of all clustering passes, with associated names (in red, based on Gorn numbering) that will be used throughout the remainder of the section. While the Two-Step Cluster Analysis helps resolve the concerns with hierarchical and non-hierarchical methods, it does not necessarily reveal the "best" discrimination between clusters. From a single pass, the cluster analysis divided the data into two clusters, but after comparing the centroids with randomly sampled individual data points, many of them were not close to the average values in each cluster. The centroids were not good indicators of the model user-pair for the clusters. Therefore, I completed a second and third pass on several of the clusters found.

For each clustering pass, the following parameters were applied:

- Method: Two-Step Analysis
- Distance measure: Log-likelihood

- Max number of clusters: 999
- Standardize variables: True
- Clustering criterion: Schwarz's Bayesian Criterion (BIC)[6]

6.3.1. Cluster descriptions

6.3.1.1. First cluster pass

The average values for each variable in the two clusters are significantly different across all the media types (Appendix A, Figure 8.1). Comparing clusters 1 and 2, the differences for every artifact interaction – in either direction – are very extreme. The distribution ratio between clusters is also high, 97.6% vs. 2.4%. The difference between clusters appears to be discrimination between highly active user pairs in cluster 2 (2.4%) versus "regular" user pairs in cluster 1 (97.65%). Chat and discussion volume is approximately thirty times greater in the high activity group, although they maintain a roughly 2-to-1 balance for chats and close to a 1-to-1 balance for discussions. Another interpretation of the numbers could be that a majority of chat associations are one-sided, where one speaker produces most of the content in a conversation. The discussion numbers for both groups are more balanced, while files do not appear to have any significant use for most user pairs.

The low number of discovered clusters (2) prevents any real generalization of what kinds of pair-wise relationships each cluster represents. There are too many instances where a single case differs drastically from the centroids of each cluster, but was probably placed there due to an even larger difference from the other centroid. Further breakdowns are necessary to discover more homogenous groups.

6.3.1.2. Second Cluster Pass (Breaking Down Group 2)

A second pass could help identify whether there is any interpretation that can be made about the cluster with high interactions. A second pass of cluster 2 finds two additional clusters. The first one, at 95.4% (cluster 2.1), closely resembles the original parent cluster (cluster 2) from the first pass, with high volume, unbalanced chats, balanced discussions, and low file usage. The smaller cluster (cluster 2.2) has even higher frequencies of associative activities, with average chats in the thousands. The ratios are similar for chats and discussions, unbalanced toward one side for chats (although less so), and more balanced in discussions. File directions were reversed.

An important point to note regarding this pass is that the discrimination levels – how well each variable helps discriminate between clusters – are no longer equal across all variables. Instead, the discussions actually take the lead role as the

[6] Used to choose between competing models (in this case, an optimal number of clusters)

discriminating component in cluster formation. Chats have a decreased role and files have very little importance in cluster discrimination.

6.3.1.3. Second Cluster Pass (Breaking Down Group 1)

Two additional clusters were identified in the second pass. Cluster 1.1, at 89.2%, closely resembles the original cluster (cluster 1) in both volume and ratio. Chats were one-sided, discussions were balanced, and file associations were insignificant. The smaller cluster 1.2, at 10.5%, had similar ratios but much higher values. Centroid values were approximately five times higher for all six variables. However, even with this increase in value, the centroid still remained lower than the numbers for the original high interaction group (cluster 2).

6.3.1.4. Third Cluster Pass (Breaking Down Cluster 1.1)

A final pass was applied to break down cluster 1.1. For the first time, three clusters were identified in the results. Cluster 1.1.1, at 77.1%, had lower chat interactions, but with a similar ratio. Discussion usage dropped to less than one, and files were negligible. Cluster 1.1.2, at 13.9%, saw very little chat use, but higher discussions (still balanced), and again, negligible files. Cluster 1.1.3, at 9%, had extremely high chat, but with all discussion and file variables less than one. It appears these user pairs only interact via the chat artifact.

As with the earlier second pass, there were drops in discrimination importance. In this case, one of the file variables dropped significantly, likely due to almost no file sharing between any of the user pairs. This was an important factor in stopping the hierarchical passes.

6.3.1.5. Summary of results

I summarize the six final clusters to be examined. These are the leaf nodes, with no follow-up clustering for any of them. The percentages are in context of the full data set, rather than from a single local cluster pass. The numbering corresponds to the labels in Figure 6.2.

2.2. (0.3% of total) = highly interactive pairs, slightly unbalanced chats, extremely high discussions, oppositely directed files as compared to chats and discussions

2.1. (2.3%) = highly interactive pairs, chats are one-sided (2-to-1), balanced discussions, files have some use, follow same directionality as chats

1.1.3. (7.8%) = only use chats, one sided (2-to-1)

1.2. (10.5%) = very active chat user pairs, one side contributes more (2-to-1), discussion use is balanced. Similar to the 67% group to a lower amount

1.1.2. (12.1%) = mostly discussion user pairs, balanced, and occasional chat

1.1.1. (67.1%) = low-level chat user pairs, where one side contributes more content (3-to-1).

6.3.2. Case Examples from the Clusters

In this section I examine case example of pairs of users in each cluster, to better understand the nature of the clusters. The case examples were selected as follows. For each case (associated pair of users), the distance measure was calculated to the average of the cluster center they belonged to, using the square root of the sum of squares between the individual vector and the centroid. The ten closest cases to the centroid average were examined. Of those ten, two representative examples were selected to explore in this section, including viewing each chat transcript the users participated in, all discussion forums where messages were posted, and each file that was shared. No details that could be used to identify any individual user are included in this presentation. All pseudonyms used are completely arbitrary.

6.3.2.1. Cluster 2.2

Cluster 2.2 contains 247 user pairs, consisting of only 4.6% of the second cluster pass and only 0.1% of the population as a whole. It is the only cluster where the average file usage is greater than one in both directions. The chat counts are extremely high, and with a 0.67-to-1 ratio, it is also the most balanced of all the clusters, as the ratio is usually at least 2-to-1 in the other clusters. There is a very high level of interaction via discussion forums, as no other cluster has averages greater than 100 in either direction. Similar to chats, the discussion ratio is also very balanced, although slightly favoring the same direction as chats.

The first example pair selected comprises Amy and Bob.

Table 6.4: Association between Amy and Bob

User1	User2	F1to2	F2to1	D1to2	D2to1	C1to2	C2to1
Cluster mean		1.29	4.77	673.95	593.48	2710.30	1820.60
Amy	Bob	1	3	103	44	3658	2495

As described in section 6.1.2, each column can be read as an associogram dependency. For example, F1to2 would mean files where User1 downloaded from User2. Based on profile information, both users are PhD students studying in the education field. Several chat transcripts were randomly selected for closer examination. I show several examples that are representative of their chat sessions. Below are portions of two examples.

Table 6.5: Parts of conversation 1 between Amy and Bob

Amy	17:57:52	N is just the number of people who responded
Bob	17:57:58	so type and carnigie - give an example of why that is important
Bob	17:58:20	correlation but between what and what
Amy	17:58:22	so
Amy	17:58:38	between whichever 2 you pick
Amy	17:59:03	go to column 2 row 1 of your output

Table 6.6: Parts of conversation 2 between Amy and Bob

Amy	14:38:12	I had never been to the venue before so that was fun
Bob	14:38:18	lovely beardy chaps arne't they?
Amy	14:38:24	the group started out great
Amy	14:38:30	yes
Bob	14:38:31	a bit shy
Amy	14:38:37	funny
Bob	14:38:47	started out?
Amy	14:38:55	had a bit of a lul in the middle

Both conversations took place inside one of the user's virtual personal offices, the first in Bob's, the latter in Amy's office. The first conversation was a discussion regarding an assignment both users had. Their discussion was balanced and informal, usually going back and forth as they collaborated on solving various problems. Occasionally, the two users would diverge and discuss off-topic issues. The second conversation was also informal, but did not have a central topic of focus. The two users were having a casual conversation, and the topics drifted around different parts of their daily lives.

Table 6.7: Parts of discussion forum contributions

Original	2006-09-24 23:24:38	… After reading it, address the following: a. Do you believe that their views adequately represent the important concerns for the future of US higher education? Expand and explain. b. What concerns, if any, should be modified -- either new issues added or others changed or deleted. Why? …
Bob	2006-10-05 08:41:04	… It just seems that a sledgehammer is being sued to crack a nut. What you may find is that individual institutions and faculty will invent ways to check the boxes while maintaining their core values - faculty are smart people! An example of this here is where staff check employability boxes by having more 'live projects'…
Amy	2006-10-10 07:39:20	…I was also struck by the commissions statement that colleges should, "'Make every effort' to calculate their 'net price', the cost of attendance once financial aid is taken into account, and share that information with consumers." I appreciate their desire to give some concessions to the US Government, but this is a ridiculous idea, given that we live

		in a capitalistic society…

Primarily, the discussion forums where both users were participants were used as part of a class-wide forum for assignments. Table 6.7 is an example of their contributions to a class-wide discussion. In most cases, an instructor initiated the creation of threads, and both users would contribute content as part of responses to inquiries related to an assignment. They tone was very formal, and while they read each other's messages, I found no direct responses between the two. The few files used were also for class-related work.

The second pair in this cluster consisted of Carly and Dan.

Table 6.8: Association between Carly and Dan

User1	User2	F1to2	F2to1	D1to2	D2to1	C1to2	C2to1
Cluster mean		1.29	4.77	673.95	593.48	2710.30	1820.60
Carly	Dan	0	8	35	46	3823	2371

This pair's interactions were a little further from the average of the cluster centers, with regard to the lower discussion interactions. However, this user pair was chosen for a particular purpose; Carly is known to be one of the longer tenured Tapped In members, and frequently assists other users on how to use the system. Therefore, it will be easier to identify this user in other clusters and compare how their interactions with other members differ (as I will do later on).

Based on profile data, one user is an education technology support consultant and the other is a middle school teacher. Here are two examples of randomly selected chat transcripts.

Table 6.9: Parts of conversation 1 between Carly and Dan

Dan	04:09:46	That is how I feel as well!
Carly	04:09:57	Which is why I'm on 18/7... moderate about 100 groups and facilitate everything I can.
Dan	04:10:21	I wish I could be here more.
Carly	04:10:21	Responsibilities imply that someone is paying for me to do these things.
Dan	04:10:31	I just get so busy with so many other things!

Table 6.10: Parts of Conversation 2 between Carly and Dan

Dan	02:50:43	We went to Applebee's and I am too full!
Dan	02:51:08	We got home 2 hours ago and I am still full!
Carly	02:51:32	You drove 120 miles to go to Applebees?
Carly	02:51:48	wonders how far you'll drive for an "In and Out Burger"?

Dan	02:51:52	I didn't. 15 minutes for me!

The first conversation took place in a public reception lobby, while the latter conversation was in Dan's virtual private office. Both were informal conversations with no focused topic. The first conversation, in a public area, was more related to the Tapped In system and could be considered work-related, although without any specific topic. The second had more personal information and revolved around day-to-day matters.

Table 6.11: Parts of discussion forum contributions

Dan	2006-03-30 18:42:13	I am a middle school computer teacher and decided I wanted to expose my students to the virtual environment. I jumped in with both feet and had a few issues. First, they chat WAY too much. I was firmer in my second class of the day, but am still concerned…
Person1	2006-03-30 20:12:00	Dan, I too teach MS students (grades 7 and 8). What kind of virtual environment are you using? A forum? Instant Message program? Wiki? Blog? I use all of the environments listed above, but I have found that it is vital to determine what the goal is of writing in a virtual environment…
Dan	2006-03-30 21:46:03	Person1, Thanks for the input. Since I am a computer teacher and we are on the computers every day, we have already had the discussions about online behavior. I have dealt with it when we have talked about internet safety and the like. Many of my 8th graders…
Carly	2006-04-02 17:53:07	Dear Person1, Thank you for your insights and suggestions regarding phpbb and nicenet... both of which are excellent solutions and both are more robust regarding threaded Discussion boards than Tapped In (I hate to say, but it's true). Another alternative to consider would be Moodle, also open source and free….

The few discussion forum interactions the two users had were limited to professional discussions regarding the Tapped In system, with no personal dialogue or any reference to their live chats. Table 6.11 is an example of their responses to a discussion about using virtual tools inside a traditional classroom, although they never directly responded to each other. Files were similarly used as notes and tools used in reference to Tapped In and other online systems.

Cluster 2.2 General Summary: Both user pairs were found in the smallest cluster of data, the 0.1 percentile. Their interactions shared many similar characteristics, with each other as well as other user pairs that were not described above. Live chat conversations were very informal, even when the users were discussing matters such as school work or using the Tapped In system. Many times, personal issues were discussed, either as the topic of the conversation, or as a digression from work-related matters. The fact that some conversations took place in a public room and others in private offices did not seem to change how the users interacted. It is possible that the users also know each other offline. Contrastingly, their interactions as mediated by the discussion forum were more formal and structured, and there was no personal information being shared. They did not use the discussion forums to extend any of their chat conversations, and only participated as part of a larger group discussion, and frequently interacted with other members of their group. File sharing showed no noticeable pattern of interaction, as it was very low in number and only used for work-related topics.

The high frequencies of chat interactions in this cluster seem to involve users who have personal relationships with each other. This cluster is the only group where chat contributes average in the thousands, which suggest long conversations. All of the examples have shown a significant amount of informal discourse between the users. Contrasting that with the formal nature between the same users in the discussion forums, it shows that the personal and professional aspects of their relationship are distributed differently across the artifacts.

6.3.2.2. Cluster 2.1

Cluster 2.1 contains 5175 user pairs, consisting of 95.4% of the second cluster pass, and 2.3% of the population as a whole. Of the first pass cluster, cluster 2.1's centroid is closer to its parent centroid (cluster 2) than cluster 2.2, with similar file, discussion, and chat interaction levels. Compared to cluster 2.2, the discussion and chat interactions are significantly less (approximately 1/6 of the values), although the file interactions are similar. The ratios are also different for chats, where the balance is almost 2-to-1. Discussions have a similar balance, although files have the opposite direction as the other cluster.

The first pair selected is Ellen and Frank.

Table 6.12: Association between Ellen and Frank

User1	User2	F1to2	F2to1	D1to2	D2to1	C1to2	C2to1
Cluster mean		4.52	0.75	91.09	74.64	531.47	281.78
Ellen	Frank	0	2	23	12	553	239

Based on user profile information, one user was a high school teacher and the other was a university instructor.

Table 6.13: Parts of conversation 1 between Ellen and Frank

Frank:	17:02:49	I know that there is a way to enlarge my screen through actions, but I don;t remember how... anyone know?
Person1:	17:02:50	I noticed that
Ellen	17:03:02	waves to Person2.

Ellen:	17:03:14	In the "Actions" drop-down menu, select "Detach."
Frank:	17:03:47	Cool, Thanks Ellen
Ellen:	17:03:48	You can then re-size the window to fit your preferences. You can also adjust the size of the text.
Frank:	17:04:04	That'
Frank:	17:04:21	That's good for these old eyes
Ellen:	17:04:22	You're very welcome. Please feel free to call me El. I had to use the whole first name simply to find a user id.

Table 6.14: Parts of conversation 2 between Ellen and Frank

Frank:	18:09:40	The up loading and downloading turnd out to be the biggest challenge. I was intrigued in the youtube - very broad-based and a little scary for high schoolers.
Frank:	18:10:06	But fun ideas also
Person3:	10:10:39	I like seeing another classroom project, Frank.
Ellen:	18:10:45	Frank, what about the music generated so much work?
Person4:	18:10:50	Frank, I was impressed by your movie and the classroom focus.
Person5:	18:10:56	me too
Frank:	18:10:59	I just couldn't find what I wanted
Ellen:	18:11:22	Ah, so it was the process of finding music rather than the process of incorporating it into the video?
Frank:	18:11:34	I had a tune in my head, but couldn't find it out there,
Person3:	18:11:43	Where did you go to find your music, Frank?
Frank:	18:11:46	So, Yes that was the problem

Both conversations took place in a room that was generally used by a group that the two users belonged to. Other members of the group were in the same room during both conversations; I have created Person# pseudonyms for members who are not Ellen or Frank (I will continue using this method in future scenarios). The group's main purpose focused on writing projects. Many of its members were teachers who were exploring technology and how to integrate it with teaching writing. They participated in "e-writing marathons" and other activities that supported their goals.

Early conversations between the two users were more formal, but as time passed, the tone grew more informal with each setting. However, the chats still maintained a professional nature and always remained on topic. Personal issues were not discussed. Chat interactions usually took place in small group settings, generally with at least four users in the room; occasionally, the two of them would be the only ones to talk, although this was rare.

Table 6.15: Parts of discussion forum contributions

Creator	2006-07-08 23:48:15	…Write a piece in response to the music website you investigated. It could be a brief "report" on the website and what you found there; be sure to include plenty of details about the site in your report. Or it could be a poem or a story or musings of your own in response to the music you heard and/or what you learned on the site. Your choice! <B>Post your writing here by clicking on the tiny <i>reply</i> link above.
Frank	2006-07-14 09:17:27	I have enjoyed reading about the wax cylinders for recording sound as a good history lesson. It is neat to hear the actual voices from long ago. I'm not so sure that my students would enjoy it as much though. They have been brought up in an age where they have seen the A&E specials that have shown it all, digital storybook style, with the same sounds the in the background…
Ellen	2006-07-17 12:39:08	I've been fascinated by podcasts lately. I use iTunes to download book-related podcasts from NPR, the New York Times, and Slate. I download Barack Obama's podcasts, simply because I find his voice his voice captivating and his perspectives intriguing…

Many discussion forums were extensions of the chat sessions. The topics were similar and the members of the group used the forums to follow up with earlier chats, or plan future ones. Table 6.15 is an example of a discussion created to discuss a writing exercise about technology and music. Both users respond to the thread several times, but never directly address each other. This also occurs in other threads. A few files were used for sharing writing samples, although it was not a frequently used artifact.

The second pair in this cluster were Hank and Greta.

Table 6.16: Association between Greta and Hank

User1	User2	F1to2	F2to1	D1to2	D2to1	C1to2	C2to1
Cluster mean		4.52	0.75	91.09	74.64	531.47	281.78
Hank	Greta	12	1	154	85	583	351

Based on user profile data, both users had interest in language arts. One was a faculty member of a university.

Table 6.17: Parts of conversation 1 between Greta and Hank

Greta:	12:32:35	Hiall
Hank:	12:32:45	Thanks for edits, Greta
Hank:	12:33:00	Person1 saved you all from reading my trashy letter.
Greta:	12:33:12	Is there a way to track changes?
Hank:	12:33:24	No, not on the whiteboard.
Person1:	12:33:31	I have the original
Greta:	12:33:32	Not a big deal

Table 6.18: Parts of conversation 2 between Greta and Hank

Hank:	15:15:02	Starting a new thread gives people a choice. That's good.
Greta:	15:15:11	Agreed
Person1:	15:15:36	But I don't think we want to pepper them with prompts
Person1:	15:15:47	So I do think this planning of who posts what is a good idea
Greta:	15:16:18	You're talking about new threads, right? Not just comments?
Person1:	15:16:34	Yes, new threads
Greta:	15:16:46	I agree
Person1:	15:17:01	Greta, have you been holding back,
Greta:	15:17:09	No
Person1:	15:17:15	that is, have you been wanting to jump on with a comment
Hank:	15:17:18	That's good.

Both conversations took place in a group room setting. This group was collaborating on designing a proposal and project for an experiment. The tone was relatively informal, although there was no discussion regarding personal matters. Chats would typically be planned in advance, with specific topics of discussion clearly prepared. All of the chats sampled were in small group settings, with at least four members present.

Table 6.19: Parts of discussion forum contributions

Hank	2006-10-11 05:25:05	…Wondering if we should be responding to them/addressing their issues yet? To keep track of them, I've made one of my obsessive/compulsive tables, and will try to monitor how/whether they get addressed (the people and the issues.) I'll post in FILES? I could attach an HTML here, but not sure how. cheers from an inservice cheerleader

Greta	2006-10-12 04:38:48	Wow! Do I have a lot of catching up to do!!!!

The two users interacted in the discussion forums mainly as an extension to their group interactions in the chat rooms. Threads were created as follow-ups to chat sessions or as a setup for planning future sessions. Unlike some of the pairs described earlier, the general tone of several discussion threads was very informal. In many cases, the messages were very short, used almost like a chat session. The example in Table 6.19 was a direct reply from Greta to Hank regarding new participants to their group. Additionally, some discussions were connected to the higher file interactions that the users had. The two users (and the group as a whole) shared files as a collaborative writing artifact. When edits or new files were introduced, the users would make discussion postings describing the changes.

Cluster 2.1 General Summary: The chat interactions all took place in small group settings and were highly interactive. There was generally an informal tone in the conversations, although no personal discussions took place. The chats were all scheduled in advance, and focused on specific goals. The groups appeared to use chats as a form of collaboration in organizing different needs, whether integrating technology in a classroom or planning an experiment. Discussions were similarly used in both cases, as a means of following up from chat discussions and planning future events. One difference to note is that while the first pair did not use files, the second pair shared files as a means of a tool to write collaborative documents, and also used the discussion forum as the primary media to discuss file updates.

6.3.2.3. Cluster 1.2

Cluster 1.2 contains 23,937 user pairs, 10.8% of the second clustering pass and 10.5% of the population as a whole. Compared to its parent cluster (cluster 1), its centroid is approximately 5-6 times higher in all artifact interactions. Although the values gives this cluster the highest average variables in the breakdown of cluster 1, it is still significantly less than any of the high interaction groups (clusters 2.1 and 2.2).

The first example pair constitutes Irina and Jack.

Table 6.20: Association between Irina and Jack

User1	User2	F1to2	F2to1	D1to2	D2to1	C1to2	C2to1
Cluster mean		0.36	0.07	22.16	18.45	104.63	51.02
Irina	Jack	0	0	23	20	102	57

The interaction data for this pair is very near the centroid values for cluster 1.2. Based on user profile information, both are high school instructors. One is also working toward a doctorate degree.

Below is a portion of a chat transcript the two users interacted in.

Table 6.21: Parts of conversation between Irina and Jack

Irina	15:59:24	I read your earlier piece about living in a small community and I like the idea of connecting to your students and being about to call with concerns. I teach in high school also. I taight for the first time a Creative Writing course and became e
Jack	16:02:12	I was in New Orleans in January 2004 as a facilitator of the New Site Leadership Institute and that's when I met Richard. We went to Cafe Du Monde at midnight for a beignet and we went to the Blue Nile to hear some jazz. Richard knows so much
Irina	16:03:11	I just read your poem about your son's wedding. Very beauriful. Thank you.
Jack	16:04:17	Oh shoot. I'm editing it right now. I wanted to have links to the poems I read aloud but it won't let me do that like I can in my blogs.
Jack	16:04:40	Or like on a regular web page
Jack	16:05:39	yes
Jack	16:08:22	Hey Irina
Irina	16:09:32	I'm working on it now. How does I post separately like you did?
Irina	16:09:43	oops how do I
Jack	16:10:08	Oh
Jack	16:10:19	Do you see it?
Jack	16:10:49	Wait
Irina	16:10:52	okay i see it
Jack	16:11:11	Do you see the folder marked June 2006 marathon?
Jack	16:11:25	hit that folder and then click on post new topic
Irina	16:11:35	i'm on the marathon part okay i think i have it
Jack	16:11:44	Great!
Irina	16:12:16	okay i'll try to get that up

Both users participated in other chats, although this instance was the only occurrence where they were in the same virtual room and frequently interacted with each other. The chat room was an artifact of a group to which they both belonged. Even though the users had never chatted before, the tone of the conversation was informal and light. There were portions where personal issues were talked about, although that could also be attributed to the topic being discussed, as the users were both reading and commenting on writings the other had created.

Table 6.22: Parts of discussion forum contributions

Creator	2006-06-07 10:51:03	… It is interesting watching writers (not the material for movies, but. . .) I wonder if teachers can learn anything just by watching their students write. Would we see what we expect? Would we change a practice?...
Irina	2006-06-07 11:50:45	… I think you can indeed learn much from watching your students write. I know I did this past year. Some were comfortable writing will sitting on the floor of a portable building. Some wandered outside to the deck for more quiet. Still others could only write with headphones and music…
Jack	2006-06-07 13:16:49	…I often HATE where I'm writing so I know my students must hate it--when writing in the classroom, it's like they think they have to write this perfunctory necessary evil paper to get the grade. That's not what I want. Your piece reminds me of advice I've gotten about writing my dissertation--keep your butt in the chair and write every day…

The users also used the discussion forums in a group setting to comment on additional writing works, contributed by themselves and other members of the group. Table 6.22 is an example of both users commenting on a writing sample from another user from their writing group, and comparing experiences in teaching and learning. Files were not used at all; instead, writing samples were pasted into discussion messages and then commented on directly in the forum.

The second pair in this cluster consisted of Kate and Leon.

Table 6.23: Association between Kate and Leon

User1	User2	F1to2	F2to1	D1to2	D2to1	C1to2	C2to1
Cluster mean		0.36	0.07	22.16	18.45	104.63	51.02
Kate	Leon	0	0	25	24	107	43

Based on user profile information, both users are teachers, working toward their master's degree in an education related field, and involved with using technology for education. Below is an example of one of their interactions from a chat transcript.

Table 6.24: Parts of conversation 1 between Kate and Leon

Leon	18:26:14	patting herself on the back
Person1	18:26:15	. o O (:))
Kate:	18:26:15	/th what does this mean?
Person2	18:26:25	. o O (Teri is a funny gal)
Leon:	18:26:9	means to only type the /th

Person6:	18:26:30	it means think
Person3:	18:26:35	im hungry
Person4:	18:26:36	yeah, but looks aren't everything Person2;)
Person6:	18:26:41	and then it automatically does your thots in a bubble
Person5:	18:26:43	type the symbol / and the the letters th

Table 6.25: Parts of conversation 2 between Kate and Leon

Person6:	18:39:57	see me?
Kate:	18:39:57	no
Leon:	18:39:59	we can see you though
Leon:	18:40:06	in the "online" category
Person3:	18:40:10	yes
Person6:	18:40:11	you can see I'm online in TI but not in the room
Person6:	18:40:14	you can't see my conversations
Person4:	18:40:16	If you yelled loud using your "mommy" voice I could hear you in another room ;)
Kate:	18:40:16	got it
Leon:	18:40:18	right
Leon:	18:40:25	it will tell you what room you are in

Both excerpts are part of a single conversation; it is the only chat session where both users had direct interactions with each other. The session took place in a group room led by a facilitator who was instructing everyone on how to use different features in Tapped In. The two users replied to each other when there were questions, but they never referred to each other's names, nor directly addressed the other person spontaneously.

Table 6.26: Parts of discussion contributions

Creator	2007-05-14 14:42:34	Here's your special place to sign in so I know you found this..having fun on your scavenger hunt?
Kate	2007-05-15 12:57:45	Here's Kate! I am really liking this environment.
Leon	2007-05-15 13:34:11	Hey, I found it!! I found it!! I am having fun, but a little nervous...transparency soon??

There were several discussion forums through which the two users interacted, although again, never directly with each other. The discussions were mainly used by the same facilitator from the chats, as a way of discussing features in Tapped In, for users to share information about themselves to the group, and discuss different collaborative efforts for technology in education. Table 6.26 is an example of a discussion thread created to practice using different features in Tapped In, and the responses from the two users regarding the exercise. Files were never used by the users.

Cluster 1.2 General Summary: There are similarities in the live chats for both user pairs. The tones were generally informal, although personal issues were not often discussed. Both pairs had very similar chat interaction numbers, approximate 100-to-50, and participated in a single session with most of their interactions. In all cases, the chats occurred inside a group chat room, and with a small group of other users (although other group members were not always talking). Unfortunately, neither user pair had any additional chat sessions after their initial encounters. It is interesting that each of the pairs have about 100 chat contributions (in one direction) as a possible indicator for a single, interactive conversation. Without additional follow up sessions, there is very little time for users to build personal relationships.

The various groups (i.e. Tapped in groups that users can be a member of) used discussion forums as a means of continuing chat sessions, organizing future events, and group interactions. None of the user pairs I examined in detail directly interacted with each other via the discussions, even though they might both post messages in the same thread. One group extended the discussion forum even further, using it as a place to view writing samples; contrasting with another group, which used the file artifacts for the same purpose.

6.3.2.4. Cluster 1.1.3

Cluster 1.1.3 contains 17769 user pairs, consisting of 9% of the third cluster pass. Compared to the population as a whole, it makes up 7.8%. It is the smallest of the three clusters that were found in the third clustering pass. The variables at the center of this cluster stand out very clearly: there are no file or discussion interactions at all. However, the chat interactions are the highest of the three clusters in this group. The ratio is also skewed in one direction, at greater than 2-to-1.

The first example pair are Mel and Nick.

Table 6.27: Association between Mel and Nick

User1	User2	F1to2	F2to1	D1to2	D2to1	C1to2	C2to1
Cluster mean		0.08	0.00	0.28	0.22	67.34	30.31
Mel	Nick	0	0	0	0	67	30

Based on user profile information, one user is listed as a college instructor and the other is a professional tutor.

Table 6.28: Parts of conversation between Mel and Nick

Mel:	19:08:18	hi Nick
Nick:	19:08:21	Hey everyone
Nick:	19:08:24	Hi Mel
Person1:	19:08:28	hi, Nick
Person2:	19:08:32	I'm here until Sunday 12:00
Nick:	19:08:33	Hey Person1
Person3	19:08:35	nods and waves to Nick
Person4:	19:08:44	Nick: I saw your web site... I'm so impressed!
Nick:	19:08:50	Person2, that's a long tapped in session
Nick:	19:08:54	Thanks Person4

I examined several chat transcripts where both users were present and contributed. However, none of them had any real form of interaction between the users. Both users belonged to the same group in Tapped In, working on professional development sessions where members were discussing lesson plans, assignments, and other education concepts. Their limited chat co-present situations were only in large group settings where the virtual rooms were used as a meeting place before the members broke up into smaller groups. This behavior repeated itself over several iterations, but the two users never ended up in the same smaller groups. They did show recognition to each other, and would often greet each other by name, but that was the extent of their chat interactions. Adhering to their cluster descriptions, they had no interactions via files or discussions. In short, these appear to be classmates who do not develop a closer relationship.

The second example pair comprises Carly and Pam. Coincidentally, they have the same interaction values as the previous pair.

Table 6.29: Association between Carly and Pam

User1	User2	F1to2	F2to1	D1to2	D2to1	C1to2	C2to1
Cluster mean		0.08	0.00	0.28	0.22	67.34	30.31
Carly	Pam	0	0	0	0	67	30

Carly has appeared in an earlier cluster example (as a high activity volunteer), and Pam is an elementary school teacher.

Table 6.30: Parts of conversation 1 between Carly and Pam

Pam:	17:08:2	Hi
Carly:	17:08:47	I'm Carly, an Education Technology Consultant and Tapped In Helpdesk volunteer. I'm in [BLOCKED].

Person1	17:09:24	wonders about our friend Person3 from Connecticut
Person1:	17:09:31	Want me to drive over to her house, Person2?
Person2:	17:10:19	Person3 is our discussion leader but she has not logged in at this time...so, we're going to choose COLLABORATION as our topic for tonight's discussion
Pam:	17:10:46	I'm from NJ. This is my third year working in the computer lab. The big push in my school this year is to help the teachers become comfortable integrating technology into their classrooms.
Person2:	17:11:04	excellent, Pam. I'm sure you'll have a lot of input for us

Table 6.31: Parts of conversation 2 between Carly and Pam

Person2:	17:15:40	my opening question was - as a library media specialist or as a tech coordinator, you ideally work closely with classroom teachers?
Pam:	17:16:12	Yes, part of my job involves working individually with classroom teachers to increase their confidence levels with technology and to help coordinate appropriate technology into classrom activities.
Person2:	17:16:32	Person4 is going to be leading a discussion for WriteTalk on an article about online resistance of teachers
Person2	17:16:42	. o O (October 13)
Person5	17:16:48	. o O (Person2read my mind)
Carly	17:16:53	wonders if he gets to attend and thwack them with a virtual mallet

The two users were co-present in the same chat room several times, but only had one occasion for interaction. The two excerpts are both part of the same conversation that took place in a public room, and was facilitated by Carly, as well as several other Tapped In leaders. Pam participated in the event, although it is not clear whether they had planned it or joined it spontaneously (as the event was broadcasted to all other chat rooms). The two users had other instances (not included in the detailed examples) of responding to each other, but never in a direct manner (i.e. never refer to each other by name). It was a small group setting, and the facilitators were able to engage the other members, and help them with questions regarding using Tapped In. Based on the direction of the interactions, Pam contributed more than twice as much as Carly, even though the session was a tech support class, where the facilitators would typically be expected to be providing much of the information. However, Carly (and the other facilitators) would often query the members about their roles in using technology in classrooms and keep them engaged, and also maintain a light overall tone and made the conversations more informal. In future chat rooms when the two users would run into each other; they would always greet each other, although no additional significant conversations took place.

Cluster 1.1.3 General Summary: Although the two example user pairs had identical interaction values, their associations with each other were very different. This is the first sign that the vector of the interaction values is not always sufficient to distinguish between different relationships. The first user pair had a superficial chat session where they briefly interacted, and left knowing each other's names, basic details, and their shared membership in a group, but nothing more. The users remained on formal terms afterwards. The second user pair was a little more interactive, as one of the users was in a facilitator role that necessitated more communication with the other members in the chat room. This approach led to a more informal tone toward the end of the chat session.

There were no files or discussions used in either pair, so the basis of their associations can only be made on the limited chat sessions that took place. The chat sessions were both in small group settings, where interactions directly between the user pairs rarely took place. However, identifying the role that Carly played offers information into the type of interaction that took place. Where the first user pair consisted of users who were peers with an equal standing, the difference in roles in the second pair helped shape the conversation that took place. The effectiveness of the facilitator in the chat session could result in a more informal relationship between the users in the future.

6.3.2.5. Cluster 1.1.2

Cluster 1.1.2 contains 27586 user pairs, consisting of 13.9% of the third clustering pass and 12.1% of the population as a whole. Of the three clusters found in the third clustering pass, it has almost no chat interactions between the user pairs. The majority of the interactions are mediated by discussion forums, which have a balanced ratio between the users. There is negligible interaction via files.

The first user pair selected consists of Rob and Sam.

Table 6.32: Association between Rob and Sam

User1	User2	F1to2	F2to1	D1to2	D2to1	C1to2	C2to1
Cluster mean		0.00	0.00	9.82	7.04	1.11	0.40
Rob	Sam	0	0	10	7	1	0

Based on profile data, Rob is a college student and user Sam is a researcher at a university.

Table 6.33: Parts of discussion forum contributions

Creator	2007-03-23 12:40:48	… Rule/ Way: A prediacte is composed of a verb or a verbal phrase that answers the question like 'what happens to the subject?' or 'what does the subject feel?' A predicate must

		have a verb in it,if nothing else….
Sam	2007-03-23 13:29:12	Thanks for the tips.
Rob	2007-03-31 10:44:52	thanks for the info, but does a predicate always need a verb?

With only a single chat contribution between the users, I did not examine any chat transcripts of the interaction. The discussion forums in which both users participated were all used as an online classroom. Threads included assignments, answers, discussions regarding assignments, hints and tips for different problems, etc. The two users never directly interacted – both users read and posted in discussions that the other user contributed content to, but they did not respond directly to the other user's messages. Rather, it appeared that they were both participants in a class and shared their assignments in the same forum. Several of the messages were quite long and detailed, and would focus on different educational concepts, although in discussions where both users participated, the majority of their messages were short (e.g. acknowledgement of thanks). Table 6.33 is an example of a discussion where both users were responding to advice on how to handle particular grammatical problems.

The second user pair consists of Tom and Uma.

Table 6.34: Association between Tom and Uma

User1	User2	F1to2	F2to1	D1to2	D2to1	C1to2	C2to1
Cluster mean		0.00	0.00	9.82	7.04	1.11	0.40
Tom	Uma	0	0	9	8	1	0

Based on profile data, both users are teachers, one at an elementary level and the other at middle school.

As before, the single chat contribution being shared does not offer any information about the users' associations, so no transcripts were observed. Both users participated in a discussion forum from a group focused on using technology in classrooms. However, their only contributions were in an introductory thread, giving some personal information. Other messages were superficial (e.g. "Thanks!" or "Agree") and did not contribute any real content. Like the previous user pair, they also did not have any direct interactions with each either; responses were made to the group as a whole or to another member.

Cluster 1.1.2 General Summary: User pairs in this cluster were not highly interactive with each other. Of the two pairs examined, the average time between joining the Tapped In network and their last activity was only two months. Even fully extended to the ten user pairs originally selected, the average remains the same. Three of the four users studied did not participate in any chat session at a significant level One of the users did join in several chat sessions, but did not contribute much content in any conversation. As with the other cluster in this pass, there was no file interaction at all.

The limited discussion interactions in this cluster suggest that the association between users was limited to only 1-2 discussion threads. Direct interactions were not made, as discussion forum associations do not require a "reply to" relationship, just that the users have accessed a forum in which the other user has posted in. Therefore, these associations are very weakly connected; this is even more evident due to the lack of any chat or file interactions between the users. The discussions were limited and did not offer much in personal information.

6.3.2.6. Cluster 1.1.1

Cluster 1.1.1 was the largest of the groups found. It consists of 152,410 user pairs, 77.1% of the third clustering pass. It is derived from the largest cluster in the first and second pass, and compared to the population as a whole, makes up 67.1%. This cluster has negligible file interactions and averages less than one discussion association in either direction. The chat interactions are also very low. Overall, these user pairs do not interact with each other to any significant degree. As the data is based on interactions between two users, and not the activity of any single user, a large cluster of user pairs that do not interact very much should be expected.

Several user pairs were examined, specifically for any chat interactions that might be useful in identifying the associations between users. However, all of the chat transcripts that were studied show no real interaction between the users. In most cases, one user would join a public room and utter messages of greeting. In a smaller percentage of cases, a person might join a room as another conversation was taking place; thus they would hear a few lines of chat before disconnecting or leaving the chat room.

The large population of this cluster should not be misinterpreted to say that most of Tapped In's users are mostly inactive or not contributing any content. First, many of the users in these pairs might be involved in high interactions with other users, and show up in the other clusters. Additionally, caution must be applied as not make any conclusions regarding individual users based on analyses using pair-wise association data. Arguably, many of these user pairs are only connected as a matter of coincidence, or did not have any reason to interact with each other. It would be difficult to make any conclusion based on information from this cluster, as any significant interactions between users would develop and change their association to better fit into one of the higher interactive clusters.

7. Results and Discussions

After the descriptions of the clusters and examples from each, I now give more detailed interpretations of the results. I start with comparing some of the different clusters and their connection to the media, followed by classifying some of the relationships that were formed by different user pairs.

7.1. Cluster Media/Interaction Comparisons

In the following section, I describe the interactions that occurred within each cluster, in the context of the media used. I compare across the different clusters, and suggest how and why the user-pairs might be interacting in varying ways.

7.1.1. Cluster Pass 1 (Clusters 1 and 2)

The first cluster pass is relatively straightforward in separating the major sub-groups of user associations. The larger group, by far, are users that do not interact much with each other. The smaller group shows significantly higher interaction in both chats and discussion forums. While the file interactions were also higher, the absolute total values are far fewer than the other two artifact types, indicating the likelihood that users in Tapped In rarely use files.

7.1.2. Cluster Pass 2 (Clusters 2.1 and 2.2)

Clusters 2.1 and 2.2 stem from a second pass of the smaller cluster found earlier. These “super user pairs” are in the minority, although their associations reveal the most information about how users can interact in Tapped In. Even though the two clusters represent a total of 2.4% of the entire set of possible pairs, there is a large difference between the clusters in the second pass as well. This indicates that the majority of super user pairs follow a typical pattern of interaction, but that a “super user+ pair” cluster also exists. In examining the chat transcripts in both clusters, there are some points that stand out. One important difference found in the sample data was that the super user+ group had more personal discussions. The users here spent a lot of time in chat sessions talking about their day-to-day lives, personal issues, and other topics with no relation to their work or Tapped In. The more “regular” group in cluster 2.2 had many conversations, but generally avoided personal discussions. It is worth noting that cluster 2.1 chats also occurred more frequently in 1-on-1 settings, with only the two users in the chat room. Cluster 2.2 chats usually occurred in small group settings. Chat sessions from both clusters supported an informal matter of conversing, although the discussions still remained focused on work rather than personal issues.

In both clusters 2.1 and 2.2, the discussion forums were used by groups the users belonged to, rather than as any form of 1-on-1 communication. This theme would repeat itself with each of the other clusters, as discussion forums became a tool for formal participation, group introductions, shared class work, reviewing projects, planning future events, or summarizing chat sessions. From an association standpoint, it is not always possible to judge any user pair based on how they interact via discussion forums. As these artifacts were being used in-group settings, one could expect the interactions to be less informal and more focused on particular topics. The significance of the high volume discussion mediated interactions could signify nothing more than the strength of a group, without giving any hint as to how any pair of users are associated. Yet it is worth noting that in these two highly interactive clusters, the users have the lowest chats-to-discussion ratio out of any of the other clusters (except for the discussion-only cluster 1.1.2).

7.1.3. Cluster 1.2 versus Clusters 2.1 and 2.2

One of the first points to note in comparing cluster 1.2 with clusters 2.1 and 2.2 is that the chat sessions do not recur. Regardless of the nature of the chats, none of the users in cluster 1.2 have repeat sessions with their counterparts. The relatively higher values of their chat interactions demonstrate a certain level of commitment in the conversation between two users, but there is almost a barrier that is not breached, preventing any deeper personal association. As such, no future chats take place. Similar to cluster 2.2, there is a generally informal tone in the chats, which occur within a small group setting. Additionally, there is very little personal information shared.

The difference in discussion-mediated interactions is also less between clusters 2.2 and 1.2, than between clusters 2.1 and 2.2. There is no major difference in how the discussions were applied; small groups were using it for organizing events, summarizing chat sessions, and asynchronous group interactions. As with the other clusters, there was little evidence of direct interaction between any of the user pairs. It follows that if groups were using discussion forums for the same purpose, the discussion-based associations between their pairs of members would be similar.

7.1.4. Compare within Cluster Pass 3 (clusters 1.1.1, 1.1.2, 1.1.3)

Clusters 1.1.2 and 1.1.3 stem from the largest clusters found in the two previous passes. In the third pass, they were in the minority when compared to cluster 1.1.1's 77.1%. However, unlike cluster 1.1.1's lack of any significant interaction, both clusters 1.1.2 and 1.1.3 showed some interaction. The major difference between the two clusters is the mediating artifact. Cluster 1.1.3's interactions are entirely chat based, while discussion forums primarily mediate cluster 1.1.2's interactions. With negligible file use by either cluster, we can better compare the use of chats and discussions using these clusters. The examples from cluster 1.1.3 demonstrate that user pairs with identical interaction patterns do not necessarily imply the same type of relationship. However, the chat values in both cases did represent a single session, with no follow up or any connection to other artifact interactions. Both chat sessions were also interactive and formal (although one session began showing signs of a more informal discussion toward the end). It can also be suggested that the number of chat contributions in both cases were not enough to form a more personal relationship; while the single session was enough for the user pairs to know enough about each other to interact in future meetings, there is no evidence of any additional chat sessions occurring.

The user pairs from cluster 1.1.2 show an even higher level of formality in their interactions. Without any chat sessions, the users could only interact via asynchronous discussions. The threaded discussions afford no real opportunity to have personal interactions, as the purposes of each discussion thread were maintained and there were no divergences. While a free flowing conversation in chat sessions, especially when facilitated by an experienced moderator, can lead to easy changes in topic, discussion forums appear to limit those occurrences. The interactions between the users in this cluster show even less personal interaction than those in cluster 1.1.3. However, classes or professional groups that wish to have a focused, formal discourse can take advantage of discussion forums for that purpose.

The common element in both cluster 1.1.2 and cluster 1.1.3 is that in both cases, the user pairs did not have future interactions with each other. Given the limited interactions that took place in either chats or discussions, that is not a surprising result. However, given that the associations that formed after interactions via the different artifacts appear to be heading toward different paths, it appears to be a situation where using the artifacts together might lead to results that are greater than the sum of its parts.

7.1.5. More About Cluster 1.1.1

Cluster 1.1.1 provides very little information regarding the associations in its set. Although it is the largest group of user pairs in both the third clustering pass (77.1%) and the population as a whole (67.1%) all of the interaction data is extremely low. There are virtually no file interactions, discussion interactions are less than one in either direction, and the chat interactions do not help us identify any type of relationship between users. The majority of the chat transcripts examined reveal that the associations are formed by either random room joins and exits, or brief encounters that consists of no more than a greeting or farewell. Whether through random passing, short-lived membership, etc., many of the members in Tapped In will frequently encounter people just as we walk by strangers on the street.

One significant direction that could benefit from the data in cluster 1.1.1 is a look at the evolution of the associated pairs. The chat transcripts reveal that most of the encounters in this cluster are random or inconsequential; however, there is a likelihood that a small percentage of these encounters could lead to developing a stronger association in the future. There is potential in many of these associations that could result in user pairs migrating to a different cluster. These opportunities could offer insight on the types of spontaneous meetings that best lead to development of stronger relationships.

7.2. User Relationships

I have examined the media comparisons across the different clusters, and how they might affect the ways in which people interact. The differences between artifact-mediated associations are visible in both the type of artifact and the direction of interaction occurring. I have also introduced some of the relationships between the user pairs, and will examine these in more detail now.

There are considerable differences in the relationships between user pairs in each of the six clusters, as defined by their interactions. The two pairs in cluster 2.2, Amy/Bob and Carly/Dan, appeared to be that of good friends. Conversations in their chat sessions were very informal, including topics outside of work or school. The users knew many details of each other's personal lives even though there is no hint that the users know each other outside of Tapped In. Even conversations that begin with work or school related topics often diverged into matters of personal nature. The balanced nature of their chats reinforces the characterization of their relationships. Furthermore, their roles, both offline and within Tapped In, did not affect how they interacted. While it might be expected that Amy and Bob have a balanced relationship because they are peers professionally, Carly and Dan have very different roles. Carly is a long tenured Tapped In member and serves as a help desk volunteer. In fact, we see this distinction in Carly's cluster 1.1.3 relationship with Pam; that chat ratio is more unbalanced than Carly's cluster 2.1 relationship with Dan, even though Dan and Pam have similar professional roles, i.e. both are K-12 teachers. Carly and Dan's cluster 2.1 relationship appears to transcend their roles, matching that of the "friend" status that Amy and Bob share.

The relationships between Ellen/Frank and Greta/Hank are very similar. Both pairs have shared in many chat sessions and discussion threads. Their chat sessions are informal and exhibit a light, friendly tone. Two other characteristics stand out: several conversations take place in a small group setting, and while the focus might occasionally digress, there is very little personal discussion. All of the chat sessions remain focused on work-related topics, as are the threaded discussions. I characterize these user pairs as peers who have worked together in professional settings (in this case, the Tapped In system) for an extended period of time. While they have become accustomed to each other, and take a more familiar tone in conversations, they still remain work colleagues, with little likelihood of personal contact outside a professional setting.

In many aspects, the user pairs from cluster 1.2, Irina/Jack and Kate/Leon, have very similar characteristics to the cluster 2.2 user pairs. Their chat sessions were informal and the tone was light, although it did not often stray toward personal topics. Most instances were in a small group setting. The users likely viewed each other as peers; this is reinforced by the fact that all of the users were teachers, although at different grade levels. The volume of chats suggests that the users only had a single, highly interactive session, without any additional follow up. These relationships appear parallel to offline relationships between professionals who do not frequently encounter each other; while the individuals appear to be on friendly terms, the relationships were never given the opportunity to develop, due to possible factors such as geographic distance. Traditionally, these users could be defined as weak ties, with the likely indicators to be high intensity of interaction, but low frequency of contact. Since all four users are teachers in different locations, I would extrapolate further to say they are boundary spanners, with the shared Tapped In chat sessions serving as the mediating boundary objects.

The user pairs in cluster 1.1.3 are the first group where chat interactions are both less than 100. While there is no basis for 100 chat contributions to serve as a limit or indicator of personal interaction levels, the chat sessions between Mel/Nick and Carly/Pam have very little direct interaction. There is no personal discussion at all, the tone is formal, and all of the chat sessions were in group settings. It does not appear that the users have any type of personal relationship with each other, although they show recognition when there is a reference to another's name. The users most likely view each other as acquaintances.

Each of the user pairs found in clusters 1.1.1 and 1.1.2 did not have enough interaction for me to classify their possible relationships. Rob/Sam and Tom/Uma, the user pairs from cluster 1.1.2, never had any direct interactions with each other in the discussion forums. With chat sessions and files contributing negligible data, it is likely the users did not know each other at all, and would likely not even recognize the other username. However, these user-pairs offer an opening to study the potential relationships that can be developed. The users in these clusters were able to still interact with each other in the virtual environment, and have the ability to continue interacting to develop stronger relationships.

In context of social relationships, I would classify cluster 2.1 pairs as friends, cluster 2.2 pairs as long-term professional peers, cluster 1.2 pairs as short-term professional peers, cluster 1.1.3 pairs as acquaintances, and cluster 1.1.1 and cluster 1.1.2 pairs as having no relationship.

Each of the cluster centroids in the study had certain characteristics that were similar: chat direction was unbalanced, discussions were balanced, and files were rarely used. The lack of file interactions could result from many factors, e.g. difficult

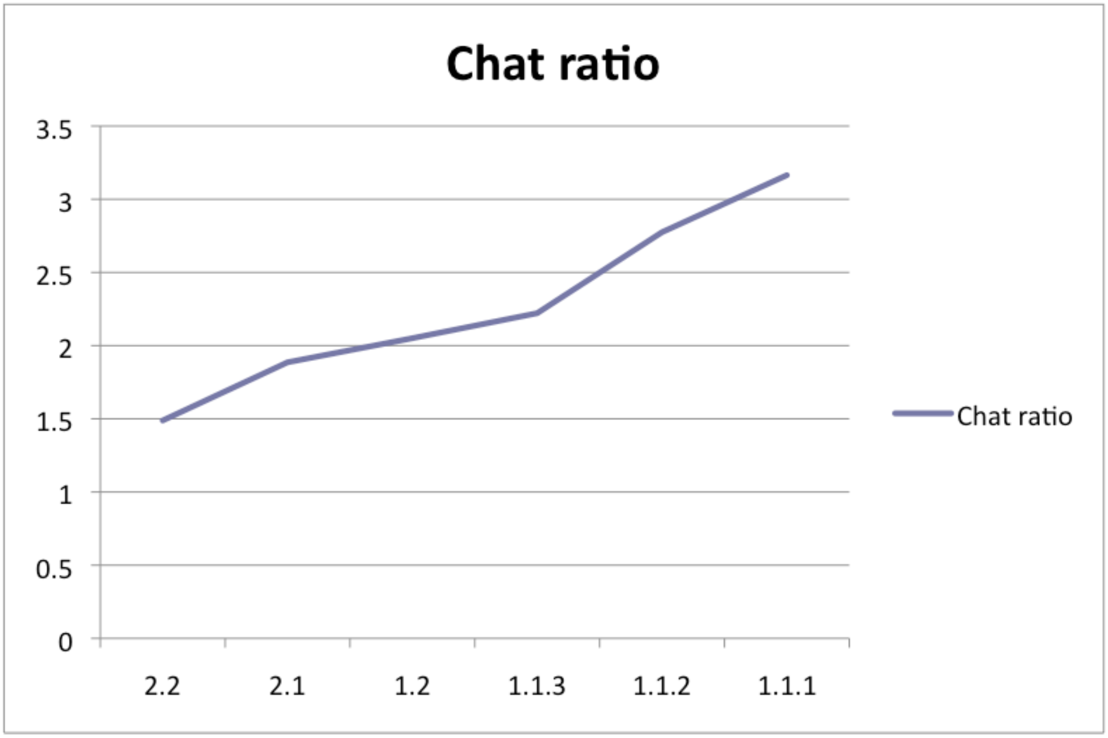

Figure 7.1: Ratio of chats between users in each of the clusters

to use interface, lack of necessity, etc. The chat imbalance, while visible in all of the centroids, had varying ratios. Figure 7.1 displays the chat ratios (Chats1to2/Chats2to1) for each of the six clusters (x-axis), ordered by decreasing volume of chat (cluster 2.2 has the highest volume). The positive slope of the curve implies that as users chat with lower frequency, there is a larger imbalance in the amount of chat contributions. This change does not appear for threaded discussions. It is likely that as user pairs increase their amount of chats, conversations become more balanced, leading to friendlier relationships. Cyclically, as user pairs steadily build their relationship, they are likely to chat more.

7.3. Limitations

Several limitations exist in this method of analysis. The first is based on the random sampling of user pairs. I selected ten cases (pair-wise associations) for each cluster that were the closest to the average center of each cluster. After examining each one, I selected two for detailed analysis and summaries based on being an average representative of the full set, with consideration for those that had detailed profiles. While these user pairs were model examples from the original ten, given the small sample size, there are many other possible relationships, especially from outliers, that will be different from the ones selected. From a dataset of 227,000 user pairs, it would be impossible to closely examine any large percentage of user interactions. However, the results that I found aligned with the patterns of interaction between the user pairs, and given the large potential for deviation, all activities fit within explainable behaviors.

A second issue with the study was the classifications of the user pairs in each cluster. I identified four separate types: friends, long-term peers, short-term peers, and acquaintances. I based my conclusions on the users' patterns of interaction and their profile information. However, no further steps were taken to validate these classifications, either through interviews, or found in any content contributed by the users to the system. Without any explicit statement from the users, the classifications I used for each user pair (and following, the clusters) remain a hypothesis. However, the analyses of the different chat sessions, both content and interaction pattern, leads me to be fairly confidence in my assessments. I believe they are strong descriptions of the actual associations that occur *within* Tapped In, without knowing how any interactions outside of the system might translate to their personal feelings for each other.

Another concern is the granularity of categories chosen for each user pair vector. Specifically, the virtual context of where the chat sessions took place was not included in the variables. Chat sessions can take place in public settings, such as the Tapped In Reception, or in private, such as user offices. Location choice can be an important factor in how users in Tapped In interact, and discriminating chat sessions by where they occurred might serve to improve the results of the clustering algorithms.

8. Conclusions

In my dissertation research, I found that how people are associated by technology reflects their relationship. Examining the various mechanisms through which people interact online provides insight into the type of relationship they have. As user relationships evolve, their interactions will also change, continuing a reflection/reaffirmation co-evolution cycle. The artifacts that people choose to use when interacting with another person, and the means by which they utilize the affordances provided, are both indicative of the relationship they share. The differences between the six clusters, and the comparisons between specific user-pair interactions, helped demonstrate that associations between users and the methods through which they interact are closely tied; that is, the descriptions of the user associations reflect the relationships that existed between them. In developing a method of research informed by many factors, including established ESDA (Sanderson & Fisher, 1994) techniques for manipulation of data, a mediation model that emphasizes the importance of artifacts and how they mediate associations between actors (Latour, 2005), application of multi-dimensional ties (Rosen & Chu, 2010), among others, I have refined a system that can be used to study those associations that exist between the macro-world of networks and the micro-world of individual interactions. However, the natural complexity of social ties shows that any description of the mediating activity does not always define a single type of relationship.

8.1. Returning to the Research Questions

Many interactions take place within Tapped In, but I chose to focus on three specific types: those mediated by files, threaded discussions, and synchronous chats. Each artifact type provides users with different affordances, which related to how they were used. While there was very little file interaction, threaded discussions and live chats were used in very different ways. The groups with the strongest relationships (clusters 1 and 2) were defined by their high volume and relatively balanced use of chat sessions. The informal tone of conversations, use of the chat rooms in 1-on-1 settings, and discussions of personal matters, were all characteristics of friends, or colleagues with long standing relationships. All of the user pairs had many interactions over the full two-year period of study (and potentially continued their relationships after). The synchronous nature of chat rooms afforded the "friendly users" a fast pace dialogue with short responses, paralleling a face-to-face conversation. In clusters where chats took place in small groups and became less balanced, I found more short-term peers and users who were merely acquaintances. There were interactions ranged from mild to intense, but were never frequent and there were few, if any, follow-ups. Threaded discussions were used in an entirely different manner. The asynchronous nature afforded more time to compose longer messages, and allowed for groups to use it without the need to coordinate members to be co-present. The users from all of the clusters posted discussion messages that were usually formal, work or school related, and without any digression to personal matters. These findings are not unknown in various media studies (Boyd & Ellison, 2007; Cox, Carr, & Hall, 2004), although I present them through a different path of analysis.

Using cluster analyses, I found six distinct groups of interaction activity patterns between pairs of Tapped In users. After selecting samples from each cluster and examining them, I found four clusters could be classified as friends, long-term professional peers, short-term professional peers, and acquaintances. The remaining two clusters did not have enough activity to be characterized in any significant manner. The clusters and user pair examples depict the relationships constituted by the users' mediated interactions. Although these results provide us with useful information, I do not claim that we are able to characterize the relationship between two users based purely on media choice. For example, the two pairs in cluster 1.1.3, Mel/Nick and Pam/Carly had identical values for both artifact use and the direction of contributions. However, their conversations were not the same, and their relationships are also likely different. Additional information is needed in order to be more accurate in classifying user relationships. For example, the study does not consider the properties of the virtual space where interactions take place. Tapped In chat rooms come in many flavors, such as private versus public. Just as we do not interact the same way at home as we do in an office, it is likely that people online would interaction differently in a private virtual space than a public room. While it is difficult to study low interaction user pairs, Bob/Amy and Carly/Dan in cluster 2.2 did not show any difference in how they interacted, whether it was in a public room or a private office. It would be interesting to know if they developed their association to allow informal conversations in public spaces, or if using private offices contributed to their familial conversations. The length of time in user pair associations was also not used here. How long any two users have known each other would help better distinguish their associations. I hope to expand this study to include more variable information, such as room properties, to help better define the full context of online interactions.

8.2. Research Significance and Discussion

Interestingly, a primary goal of social network analysis is the study of social relationships, that is, the study of ties in a social network, but neither network-centric nor ego-centric methods focus on ties (even the terms direct attention to the network and the ego rather than the ties). In a sense, both network-centric and ego-centric views can be used to study ties. Egocentric degree measurements can tell us how many ties a node has; network-centric density analysis gives us an idea of how connected a network is by comparing existing ties with unrealized ties. However, degree measurements tell us aggregate information, and in the context of a node; density is a computation associated with the network as a whole. No method or metric informs us about a single tie, other than strength, which can be misleading (Marsden & Campbell, 1984). In contrast, I focus on dyadic ties as the most important unit to be studied. The tie-centric view of a network can give us a better understanding of the associations that people can have with each other by providing a view of the different variables that constitute a network tie. Breaking down a tie into its constituent components, whether they are socio-metric data (e.g. duration of relationship) or mediating technical artifacts (e.g. live chat), allows us to study the ties themselves. In characterizing the ties, especially in a multi-dimensional context, I am describing the sociotechnical system in terms of distributions across the digital artifacts used to mediate the interactions between actors.

Inherent in this line of study is that we are no longer situating the research in either a macro-level view or micro-level view. Instead, it is in this meso-level, the “neglected scale in the socialization process” (Licoppe & Smoreda, 2005), where we conduct our studies. In SNA, one can quickly jump from a network-centric macro-level view directly into a single node in an egocentric micro-level view. By focusing on ties, we are viewing data as constructed by exactly two nodes. From this point, we can aggregate tie data (as we do with the various clustering analyses) or break down individual ties into the variables that describe it (e.g. to study multi-mediated relationships). The research is attempting to connect user associations with multi-mediated relationships. This would not be possible with either a network-centric or egocentric view of data.

The associogram contributed many of the insights and results in this study. In focusing on the meso-level view, we developed a mediation model to analyze the relationship between the social and the technical. Within the associogram, we can still conduct different levels of analysis: identify how actors and artifacts are situated in context of the entire network (Suthers & Chu, 2012), or study a single dyadic pair of users to find patterns of interaction as described by mediating artifacts. Associograms also provide information on the direction in which contributions from interactions flow, and, using an actant view of actors and artifacts, can help us to understand how artifacts can “participate” in a sociotechnical system. The clustering analysis showed that users in a sociotechnical system can be grouped by their interactions with other members, and information about their relationships are derivable from these patterns. In studying how interactions are mediated, we are giving a richer description of the relationships that people have, and an understanding of how those relationships can co-evolve with the artifact-mediated interactions. These descriptions can also provide a basis of artifact-mediated associations for future studies on sociotechnical system design.

This method of research not only provides a new lens with which to analyze data, but also acts as a bridge in connecting the forest with the trees. Breaking network-level data into clusters of ties can help step us toward what individual nodes might represent. In the earlier cluster analyses, I briefly mention roles that individuals play, whether it is a member of an organization or facilitator of a chat session. These can be studied at different levels, as individuals with different roles in a socio-technical network can be structurally situated in particular locations (Cross & Prusak, 2002; Gleave, Welser, Lento, & Smith, 2009; P. Williams, 2002), or have distinct egocentric characteristics. But the ability to view people as part of many different dyadic associations can help us better understand why they might interact differently across different sub-groups or individuals. This can offer an alternative perspective to traditional SNA concepts, e.g. determining roles via structural properties. While two actors might have identical egocentric networks in a sociogram, as defined by structure and other SNA metrics, their ties with their neighbors might be based on different dyadic associations. For example, a project manager and a third party consultant might have the same network ties to a project group's members through online interactions. However, examining their associations, as defined by the means through which they interact, can inform us as to the varying levels of collaborative work. The difference between a manager using weekly group meetings, when compared to the consultant meeting with individual members 1-on-1 daily, might give us clues as to the real roles that exist. The pair-wise associations between the consultant and the members of the team will be significantly different than those between the manager and the members. We can study the roles people have by identifying their different associations, how these associations might differ (or not) between different actors, and the implications in how they define and change a person's role.

This research can potentially offer many recommendations to sociotechnical system design. The clustering analysis identified potential boundary spanners that have a strong potential in resource sharing and dissemination of information. If we further traced a longitudinal view of the associations between the users, it can provide hints as to how the users navigate between different groups, a view of how users developed into roles, and when information is shared. The study also showed differences between users with personal relationships and those with professional relationships. How the users organized their chat sessions (e.g. how spontaneous sessions began, and the differences between the 1-on1 and small group sessions) is an important aspect in designing features of a synchronous interaction medium. The ability of a single artifact to afford different forms of interactions can potentially motivate actors to interact more frequently, and with better results.

This exploratory study was able to successfully answer its original research questions, but it also revealed other questions to be addressed. A mediation model view using associograms, combined with clustering analyses helped expose how relationships are mediated. However, the study also revealed additional gaps in theory and research methodology. Several examples demonstrated that patterns of interaction do not necessarily provide irrefutable evidence of certain relationships. Interaction patterns that appear alike, or even identical, do not always mean the social relationships between users will be the same. The method used to form associations is valid, and the associogram is a sound model to study mediated interactions, but there are other variables – besides mediating artifacts – that need to be considered when describing the complex social relationships between actors. For example, we can learn additional information about roles in user-pairs within chat contexts by analyzing who is responding to whom (Suthers & Desiato, 2012).

There are further analyses that can be done with this data. I have already mentioned the possibilities of including more variables for better clustering, and temporal data to study the evolution of user-pair associations. However, this approach could be used for analyses in other settings. The mediation model focused on how artifacts mediate interactions between actors. However, by swapping actor with artifact, we can also apply the same methods in examining how actors affect the mediation of artifacts. There can exist many actors with associations between similar artifacts, and we can learn many things about both actor and artifact by examining the patterns with which *artifact-pairs* have, as mediated by actors. The same information we discovered about user associations and possible roles can be extended to the artifacts in sociotechnical systems. Through frameworks which emphasize the agency of non-human artifacts (Hutchins, 1995; Latour, 2005), we can begin exploring the possibility of an artifact-centric analysis, in context of actor mediations.

8.3. Future directions

There are several steps I will take to take in follow up analyses with this research. I plan to include two variables that were not included in my cluster analyses. The first is a finer granularity of virtual space categories, such as comparing interactions that occur in public spaces versus private offices, or 1-on-1 conversations and small group settings. In some cases, such as with cluster 2.2, there did not appear to be major differences in the examples, but other user-pairs, as in cluster 1.1.3, showed hints where differences might be visible. A second variable that can be immediately examined is the temporal component for associations. How long users have known each other, the length of each interaction, and the "age" of users in the system can give more precise reasons of why user-pairs exist in the clusters they were found in. Inherent in this focus is how user-pairs might migrate from one cluster to another, and the mediating factors that artifacts have on this process.

A more ambitious step is to apply the analysis to a different community, either through an existing system such as disCourse or new ones that might use more Web 2.0 artifacts. Tapped In offered a large amount of user association data that readily supported my analysis, with a range of artifacts with which I could discriminate between how users interacted. However, the results might not be immediately applied in a generalized manner; while the artifacts might share many affordances as those in other sociotechnical systems, advances in technology have introduced many new ways by which people interact online. The EEC framework and associogram model (Suthers & Rosen, 2011) are not system-specific, and can be applied to any system for study.

9. Appendix A

This section provides more detailed results of the different clusters that were found using SPSS. Additional data includes the relative sizes of each cluster in context of a single pass of the data, some statistical information, and the varying levels of input importance for each variable. SPSS determines the level at which each variable contributes to the discrimination of the clusters, and identifies this as input importance. In the following figures, darkly shaded boxes are variables with high input importance, and boxes with no shading are variables with no input importance in the cluster.

9.1. First cluster algorithm pass of full dataset

This is the first clustering pass of the full dataset. The resulting clusters are labeled cluster 1 and cluster 2. Note that all variables have maximum (darkly shaded) input importance, each contributing equally to the cluster discrimination.

Name	Cluster 1	Cluster 2
Size	97.6% (221702)	2.4% (5422)
Variables	Chats1to2 21.60	Chats1to2 630.73
	Chats2to1 9.50	Chats2to1 351.88
	Discussions1to2 4.05	Discussions1to2 117.64
	Discussions2to1 3.08	Discussions2to1 98.27
	Files1to2 0.05	Files1to2 4.37
	Files2to1 0.01	Files2to1 0.94

Figure 9.1: First cluster algorithm pass

Size of smallest cluster	5422
Size of largest cluster	221702
Ratio of largest to smallest clusters	40.89

Figure 9.2: Statistics of first cluster algorithm pass

9.2. *Second cluster algorithm pass of Cluster 1*

This is the second clustering pass, using the dataset from cluster 1. The resulting clusters are labeled cluster 1.1 and cluster 1.2. Again, all variables have maximum input importance.

Name	Cluster 1.1	Cluster 1.2
Size	89.2% (197765)	10.8% (23937)
Variables	Chats1to2 11.55	Chats1to2 104.63
	Chats2to1 4.47	Chats2to1 51.02
	Discussions1to2 1.86	Discussions1to2 22.16
	Discussions2to1 1.22	Discussions2to1 18.45
	Files1to2 0.01	Files1to2 0.36
	Files2to1 0.00	Files2to1 0.07

Figure 9.3: Second cluster algorithm pass, from Cluster 1.

Size of smallest cluster	23937
Size of largest cluster	197765
Ratio of largest to smallest clusters	8.26

Figure 9.4: Statistics of second cluster algorithm pass

9.3. Second cluster algorithm pass of cluster 2

This is the second clustering pass, using the dataset from cluster 2. The resulting clusters are labeled cluster 2.1 and cluster 2.2. In this instance, there are three varying levels of input importance. Discussions and chats contribute more to the cluster discrimination, with discussions having a slightly higher weight. Files were significantly lower than the other two artifacts.

Name	Cluster 2.1	Cluster 2.2
Size	95.4% (5175)	4.6% (247)
Variables	Chats1to2 531.47	Chats1to2 2710.30
	Chats2to1 281.78	Chats2to1 1820.60
	Discussions1to2 91.09	Discussions1to2 673.95
	Discussions2to1 74.64	Discussions2to1 593.48
	Files1to2 4.52	Files1to2 1.29
	Files2to1 0.75	Files2to1 4.77

Figure 9.5: Second cluster algorithm pass of Cluster 2

Size of smallest cluster	247
Size of largest cluster	5175
Ratio of largest to smallest clusters	20.95

Figure 9.6: Statistics of second cluster algorithm pass

9.4. Third cluster algorithm pass of cluster 1.1

This is the third clustering pass, using the dataset from cluster 1.1. The resulting clusters are labeled cluster 1.1.1, cluster 1.1.2, and cluster 1.1.3. In this instance, all variables have maximum input importance except for Files2to1. The average found in all three clusters is the same, at 0.00, which does not provide any discrimination factor in separating the clusters.

Name	Cluster 1.1.1	Cluster 1.1.2	Cluster 1.1.3
Size	77.1% (152410)	13.9% (27586)	9.0% (17769)
Variables	Chats1to2 6.93	Chats1to2 1.11	Chats1to2 67.34
	Chats2to1 2.19	Chats2to1 0.40	Chats2to1 30.31
	Discussions1to2 0.60	Discussions1to2 9.82	Discussions1to2 0.28
	Discussions2to1 0.29	Discussions2to1 7.04	Discussions2to1 0.22
	Files1to2 0.00	Files1to2 0.00	Files1to2 0.08
	Files2to1 0.00	Files2to1 0.00	Files2to1 0.00

Figure 9.7: Third cluster algorithm pass, from Cluster 1.1

Size of smallest cluster	17769
Size of largest cluster	152410
Ratio of largest to smallest clusters	8.58

Figure 9.8: Statistics of third cluster algorithm pass

10. References

Alba, R., & Kadushin, C. (1976). The Intersection of Social Circles: A New Measure of Social Proximity in Networks. *Sociological Methods Research, 5*(1), 77-102.

Aldrich, H., & Herker, D. (1977). Boundary Spanning Roles and Organization Structure. *The Academy of Management Review, 2*(2), 217-230.

Barnett, G. (2001). A longitudinal analysis of the international telecommunications network: 1978-1996. *American Behavior Scientist, 44*(10), 1638-1655.

Beer, D., & Burrows, R. (2007). Sociology and, of and in Web 2.0: Some Initial Considerations. *Sociological Research Online, 12*(5).

Berger-Wolf, T., & Saia, J. (2006). *A framework for analysis of dynamic social networks.* Paper presented at the KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining.

Borgatti, S., Jones, C., & Everett, M. (1998). Network Measures of Social Capital. *Connections, 21*(2), 27-36.

Boyd, D., & Ellison, N. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210-230.

Breiger, R. (1974). The Duality of Persons and Groups. *Social Forces, 53*(2), 181-190.

Burt, R. (1995). *Structural Holes: The Social Structure of Competition*: Harvard University Press.

Burt, R. (2000). The Network Structure of Social Capital. *Research in Organizational Behavior, 22*, 345-423.

Burt, R. (2001). Structural Holes versus Network Closure as Social Capital. In R. S. Burt, N. Lin & K. Cook (Eds.), *Social Capital: Theory and Research* (pp. 31-56): Aldine de Gruyter.

Callon, M. (1986). The sociology of an actor-network: The case of the electric vehicle. In M. Callon, J. Law & A. Rip (Eds.), *Mapping the Dynamics of Science and Technology*: Macmillan Press.

Carley, K. (2003). Dynamic Network Analysis. In K. Carley, R. Breiger & P. Pattison (Eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers* (pp. 133-145). Washington, DC: The Nat. Acad. Press.

Carley, K., Dombroski, M., Tsvetovat, M., Reminga, J., & Kamneve, N. (2003). *Destabilizing Dynamic Covert Networks.* Paper presented at the Proceedings of the 8 International Command and Control Research and Technology Symposium.

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

Chen, C. (1999). Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing and Management, 35*(2), 401-420.

Chu, K.-H., Rosen, D., & Suthers, D. D. (2012). *Multi-Modal Multi-Granular Analysis: Exploring Communication Networks Under Varying Lenses.* Paper presented at the Hawaii International Conference on the System Sciences (HICSS45), Grand Wailea, Maui, Hawaii.

Chu, K.-H., Suthers, D. D., & Rosen, D. (2012). *Uncovering associations in a sociotechnical network.* Paper presented at the Hawaii International Conference on the System Sciences (HICSS45), Grand Wailea, Maui, Hawaii.
Coleman, J. (1988). Social Capital in the Creation of Human Capital. *The American Journal of Sociology, 94*, S95-S120.
Cox, G., Carr, T., & Hall, M. (2004). Evaluating the use of synchronous communication in two blended courses. *Journal of Computer Assisted Learning, 20*, 183-193.
Cross, R., & Parker, A. (2004). *The Hidden Power of Social Networks: Understanding How Work Really Gets Done in Organizations*: {Harvard Business School Press}.
Cross, R., Parker, A., & Borgatti, S. (2000). A bird's-eye view: Using social network analysis to improve knowledge creation and sharing. *Knowledge Directions, 2*(1), 48-61.
Cross, R., & Prusak, L. (2002). The people who make organizations go--or stop. *Harvard business review, 80*(6).
Davis, A., Gardner, B. B., & Gardner, M. R. (1941). *Deep South: A social anthropological study of caste and class*. Chicago: The University of Chicago Press.
de Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory Social Network Analysis with Pajek (Structural Analysis in the Social Sciences)*: {Cambridge University Press}.
Derry, S., & Fischer, G. (2005). *Toward a Model and Theory for Transdisciplinary Graduate Education.* Paper presented at the American Educational Research Association, Montreal.
Donath, J. S., & Boyd, D. (2004). Public displays of connection. *BT Technology Journal, 22*, 71-82.
Ellison, N., Steinfield, C., & Lampe, C. (2007). The Benefits of Facebook "Friends:" Social Capital and College Students Use of Online Social Network Sites. *Journal of Computer-Mediated Communication, 12*(4), 1143-1168.
Farooq, U., Schank, P., Harris, A., Fusco, J., & Schlager, M. (2007). Sustaining a community computing infrastructure for online teacher professional development: A Case Study of Designing Tapped In. *Computer Supported Cooperative Work, 16*, 397-429.
Fisher, C., & Sanderson, P. (1996). Exploratory sequential data analysis: exploring continuous observational data. *interactions, 3*(2), 25-34.
Friedkin, N. (1980). A test of structural features of granovetter's strength of weak ties theory. *Social Networks, 2*(4), 411-422.
Galston, W. (2000). Does the Internet Strengthen Community? *National Civic Review, 89*(3), 193-202.
Garton, L., Haythornthwaite, C., & Wellman, B. (1997). Studying Online Social Networks. *Journal of Computer-Mediated Communication, 3*(1), 0-0.
Gibson, J. J. (1977). The theory of affordances. In R. E. S. J. Bransford (Ed.), *Perceiving, Acting, and Knowing*. Hillsdale, NJ: Lawrence Erlbaum Associations.

Gleave, E., Welser, H., Lento, T., & Smith, M. (2009). *A Conceptual and Operational Definition of ‚ÄòSocial Role‚Äô in Online Community.* Paper presented at the 42nd Hawaii International Conference on System Sciences, Waikoloa, HI.

Gorn, S. (1967). Explicit Definitions and Linguistic Dominoes. In J. H. S. Takasu (Ed.), *Systems and Computer Science* (pp. 77-115): University of Toronto Press.

Granovetter, M. (1973). The Strength of Weak Ties. *The American Journal of Sociology, 78*(6), 1360-1380.

Hollan, J., & Stornetta, S. (1992). *Beyond being there.* Paper presented at the SIGCHI conference on Human factors in computing systems (CHI 92), Monterey, California, USA.

Hutchins, E. (1995). *Cognition in the Wild*: MIT Press.

Joseph, S., Lid, V., & Suthers, D. D. (2007). *Transcendent Communities.* Paper presented at the The Computer Supported Collaborative Learning (CSCL) Conference 2007, New Brunswick, NJ.

Kling, R. (2007). What Is Social Informatics and Why Does It Matter? *The Information Society, 23*(4), 205-220.

Kollock, P. (1998). The economies of online cooperation: gifts and public goods in cyberspace. In *Communities in the cyberspace* (pp. 259-262). London: Routledge.

Kossinets, G., & Watts, D. (2006). Empirical Analysis of an Evolving Social Network. *Science, 311*(5757), 88-90.

Krackhardt, D. (1992). The Strength of Strong Ties: The Importance of Philos in Organizations. In N. Nohria & R. Eccles (Eds.), *Networks and Organizations: Structure, Form,and Action* (pp. 216-239): Harvard Business School Press.

Latour, B. (2005). *Reassembling the Social: An Introduction to Actor-network-theory*: Oxford University Press.

Levina, N., & Vaast, E. (2005). The Emergence of Boundary Spanning Competence in Practice: Implications for Implementation and Use of Information Systems. *MIS Quarterly, 29*(2), 335-363.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks, 30*(4), 330-342.

Licoppe, C., & Smoreda, Z. (2005). Are social networks technologically embedded?: How networks are changing today with changes in communication technology. *Social Networks, 27*(4), 317-335.

Lin, N. (1999). Building a Network Theory of Social Capital. *Connections, 22*(1), 28-51.

Lin, N., Dayton, P., & Greenwald, P. (1978). Analyzing the Instrumental Use of Relations in the Context of Social Structure. *Sociological Methods Research, 7*(2), 149-166.

Lin, N., Vaughn, J., & Ensel, W. (1981). Social Resources and Occupational Status Attainment. *Social Forces, 59*(4), 1163-1181.

Lloyd, S. P. (1982). Least squares quanization in PCM. *IEEE Transactions on Information Theory, 28*(2), 129-137.

Marsden, P., & Campbell, K. (1984). Measuring Tie Strength. *Social Forces, 63*(2), 482-501.

Norman, D. (1988). *The psychology of everyday things*. New York: Doubleday.

Norman, D. (1999). Affordance, conventions, and design. *Interactions, 6*(3), 38-43.

Norman, D. (2002). *The Design of Everyday Things*: Basic Books.

Norman, D. (2008). THE WAY I SEE IT: Signifiers, not affordances. *interactions, 15*(6), 18-19.

Olsen, G. M., & Olsen, J. S. (2000). Distance matters. *Human-Computer Interaction, 15*(2), 139-178.

Petroczi, A., Nepusz, T., & Bazso, F. (2007). Measuring tie-strength in virtual social networks. *Connections, 27*(2), 39-52.

Pinch, T., & Bijker, W. (1984). 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), 399-441.

Putnam, R. (2000). *Bowling Alone: The Collapse and Revival of American Community*: {Simon & Schuster}.

Resnick, P. (2002). Beyond Bowling Together: SocioTechnical Capital. In J. M. Carroll (Ed.), *HCI in the new millenium* (pp. 247-272): Addison-Wesley.

Ridings, C., & Gefen, D. (2004). Virtual Community Attraction: Why People Hang Out Online. *Journal of Computer-Mediated Communication, 10*(1), 00-00.

Rogers, E. (2003). *Diffusion of Innovations, 5th Edition*: {Free Press}.

Rosen, D., & Chu, K.-H. (2010). *The utility of ties: A reconceptualization and computational exploration of social network tie measures*. Paper presented at the International Network of Social Network Analysis Sunbelt Conference.

Rosen, D., & Chu, K.-H. (2011). The Utility of Communication Network Ties: Reconceptualizing the Social Network Tie Measure. In *Proceedings of the Hawaii International Conference on the System Sciences (HICSS-44), January 4-7, 2011, Kauai, Hawaii*.

Sanderson, P., & Fisher, C. (1994). Exploratory Sequential Data Analysis: Foundations. *Human-Computer Interaction, 9*(3), 251-317.

Schlager, M., Fusco, J., & Schank, P. (2002). Evolution of an Online Education Community of Practice. In K. Renninger & W. Shumar (Eds.), *Building Virtual Communities* (pp. 129-158): Cambridge University Press.

Shi, X., Adamic, L., & Strauss, M. (2007). Networks of strong ties. *Physica A: Statistical Mechanics and its Applications, 378*(1), 33-47.

Star, S., & Griesemer, J. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*(3), 387-420.

Suthers, D. D., & Chu, K.-H. (2012). *Multi-mediated community structure in a socio-technical network*. Paper presented at the Learning Analytics and Knowledge (LAK12).

Suthers, D. D., Chu, K.-H., & Joseph, S. (2009). *Bridging Socio-Technical Capital in an Online Learning Environment.* Paper presented at the 42nd Hawaii International Conference on the System Sciences (HICSS-42), Waikoloa, HI.

Suthers, D. D., & Desiato, C. (2012). *Exposing chat features through analysis of uptake between contributions.* Paper presented at the Hawaii International Conference on the System Sciences (HICSS45), Grand Wailea, Maui, Hawaii.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2007). *A framework for eclectic analysis of collaborative interaction.* Paper presented at the The Computer Supported Collaborative Learning (CSCL) Conference 2007.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer Supported Collaborative Learning, 5*(1), 5-42.

Suthers, D. D., & Rosen, D. (2011). *A unified framework for multi-level analysis of distributed learning*. Paper presented at the First International Conference on Learning Analytics and Knowledge.

Tsai, W., & Ghoshal, S. (1998). Social Capital and Value Creation: The Role of Intrafirm Networks. *The Academy of Management Journal, 41*(4), 464-476.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis : Methods and Applications (Structural Analysis in the Social Sciences)*: Cambridge University Press.

Wellman, B. (1982). Studying personal communities. In P. Marsden & N. Lin (Eds.), *Social structure and network analysis* (pp. 66-81). Beverly Hills, CA: Sage.

Wellman, B., & Gulia, M. (1998). Virtual communities as communities: Net surfers don't ride alone. In *Communities in Cyberspace*. London: Routledge.

Wellman, B., Haase, A., Boase, J., Chen, W., Hampton, K., Diaz, I., et al. (2003). The Social Affordances of the Internet for Networked Individualism. *Journal of Computer-Mediated Communication, 8*(3), 0-0.

Wellman, B., & Wortley, S. (1990). Different Strokes from Different Folks: Community Ties and Social Support. *The American Journal of Sociology, 96*(3), 558-588.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning and Identity*. Cambridge: Cambridge University Press.

Williams, D. (2006). On and Off the 'Net: Scales for Social Capital in an Online Era. *Journal of Computer-Mediated Communication, 11*(2), 593-628.

Williams, P. (2002). The Competent Boundary Spanner. *Public Administration*, 103-124.