ABSTRACT

The Network as Dependent Variable: Antecedents of Online Social Network Formation

Wallace Chipidza, Ph.D.

Mentor: John F. Tripp, Ph.D.

Although social networks change over time, network dynamics remain an understudied area of social network research. The detailed logs that are a feature of online social networks allow for easier observation and capturing of social network evolution. This study examines the evolution of an online social network. The network is proposed to evolve with varying tendencies of density, reciprocity, centralization, and demographic influences under two competing theories – social capital theory and social comparison theory. Using a unique longitudinal data set of email communications that span a period of twenty-six months, the study examines the dynamics of the email communication network of a technology company. Prior social network research is dominated by descriptive studies. This study is different because it not only tests the statistical significance of the influence of structural and demographic factors on network evolution, but also their relative strengths. Further, the study also examines how online social networks differ from offline social networks. It examines the differences between the (online) email communication network and the (offline) advice, brainstorming, and working networks within the same organization. The study finds that the email network

evolves with tendencies towards high density, high centralization, and high reciprocity. Demographic influences are inconsistent over time; they are salient only in certain months. The study also finds that online networks exhibit higher levels of density, centralization, and reciprocity compared to offline networks. As such, the study contributes to theory by finding (I) antecedents of social network formation and (II) differences between online and offline networks.

The Network as Dependent Variable: Antecedents of Online Social Network Formation

by

Wallace Chipidza, B.B.S., M.S.

A Dissertation

Approved by the Department of Information Systems

_____

Jonathan Trower, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Doctor of Philosophy

Approved by the Dissertation Committee

_____

John F. Tripp, Ph.D., Chairperson

_____

Dorothy Leidner, Ph.D.

_____

Gina Green, Ph.D.

_____

Stacie Petter, Ph.D.

_____

Nicholas Benson, Ph.D.

Accepted by the Graduate School

May 2018

_____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This dissertation would not have been possible without the exceptional supervision of Dr. John Tripp. I am immensely thankful for his skillful guidance and jovial but comprehensive supervision throughout this process. I am especially pleased to be his first PhD student; the first of a lucky many I'm sure. I am also extremely thankful to Dr. Dorothy Leidner who has made my stay at Baylor University so memorable. I have learned a lot from her in my dissertation work and our other research collaborations; she has made me dare to imagine, and her lessons in creativity will not be forgotten.

Dr. Stacie Petter's help was instrumental in improving this manuscript. I am thankful for her wise advice, which invariably simplified my work. Her advice has not been limited to my dissertation, and I emerge wiser from every meeting with her. I'm also extremely grateful to Dr. Gina Green, who was my first research collaborator. She introduced me to the world of academic publication. Her insightful and detailed feedback on this dissertation helped me clarify numerous blind spots. I'm also extremely thankful to Dr. Nicholas Benson for devoting so much time and energy in improving this manuscript. His suggestions early on helped me expand the utility of this dissertation, and I'm confident that future publications arising from this study will draw specifically on many of his astute observations.

I'm also thankful to Baylor's Information Systems faculty at large. I thank Dr. Kayworth and Dr. Trower, for supporting my research training in the position of department chair and for general career advice. I thank my other academic collaborators

Dr. Burleson, Dr. Koch, and Dr. Riemenschneider for teaching me the ropes of qualitative and quantitative research in IS. I thank Dr. Wakefield, who taught me my first research class at Baylor; the influence of that class still echoes in the quantitative work that I do.

I'm thankful towards my academic siblings, those who graduated before me and those that enrolled after me. Drs. Giddens, Yan, Milic, Balozian, and Zhang taught me how to navigate the program, and I had fun learning from them. I thank my friends Jordana George (who I took so many classes with) and Eric Villafranca; discussions with them are always so interesting. I'm thankful to James, Uchenna, Khalid, and Pete for making me feel important by asking useful questions about the program.

I'm also thankful to my family and friends for providing me with social support throughout my educational journey. I especially thank my late father, who had more faith in me than I ever had in myself. I thank the whole Chipidza family – Mavis, Kevin, Tafadzwa, Nyasha, Fallon (who always manages to make me laugh), Tino, Farai, Denise, and Darrell. Deep gratitude is due to the smart and beautiful Tendai Gwanzura, who has been my staunchest cheerleader during this process. Lastly I thank the friends that have motivated me to do the best that I can: Wisdom Nyaguwa, John Mhuriro, Linroy Munetsi, Kuda Mutombo, Johnson Mazombe, Gift Ntuli, Tapiwa Musasa, George Chirenda, Tashinga Pambayi, and Chalenge Masekera.

DEDICATION

To my late parents Shepherd Chipidza and Nyadzani Jambaya
Memories of you inspire me to reach higher

CHAPTER ONE

Introduction

Social networks are structures that arise out of social interactions, a large proportion of which are increasingly being mediated by electronic tools. In fact, electronic communication now far outpaces face to face communication within organizations (Dabbish et al. 2005; Gloor 2016; Putzke et al. 2010). Online networks – social networks arising out of these electronically mediated interactions – increasingly attract practitioner and scholarly interest (e.g. Putzke et al. 2010; Sykes, Venkatesh, and Gosain 2009; Zhang and Venkatesh 2013). One particularly important instance of an online network is the network that arises out of email communication. Consistent with broader social network research, email networks better explain organizational outcomes than formal structures of the organization (Gloor 2016; Krackhardt and Hanson 1993); as such, email networks are consequential structures. For example, influential members in email networks tend to perform better, register high client satisfaction, and express high subjective well-being relative to others (Aral and Van Alstyne 2011; Diesner and Carley 2005; Gloor 2016). At the same time, email network position is also related to multiple negative outcomes like information overload and decreased job satisfaction (Dabbish et al. 2005; Merten and Gloor 2010). The outcomes associated with online networks are thus somewhat understood, but the factors that cause their formation are not; these factors are even less understood for organizational social networks (Borgatti et al. 2009a; Johnson, Faraj, and Kudaravalli 2014).

Generally, studies on organizational social networks proceed in the following manner. The researcher surveys participants on whom they are connected to by some criterion e.g. for an advice network, participants are asked how many times they sought advice on work-related tasks from every other employee (see Cross et al. 2001; Zhang and Venkatesh 2013). The network is then drawn using the resulting self-reported information. Such studies are therefore subject to self-report bias; for example, in a friendship network, participants may inaccurately report that they are friendly with the most popular people in the network. More importantly, subsequent calculations are then done on this cross-sectional snapshot of the network.

However, social networks are not static - they change over time; relationships among participants may strengthen and persist, or weaken and even terminate over time (Snijders, Van de Bunt, and Steglich 2010). But social network evolution has received little attention in the literature because it is difficult to observe in the natural world (Borgatti et al. 2009b; Schaefer et al. 2010). For example, to observe the evolution of a friendship network, the researcher would have to monitor multiple relationships over an extended period. Even then, the researcher might miss certain network events; some friendships might die and resurrect between periods of observation. Hence, studying network evolution avails finer explanations for observed network phenomena. This study thus examines how structural and demographic factors determine social network formation, with particular focus on online networks. By examining network evolution, we can answer several questions, for example (I) how individuals or groups become isolated from their peers over time, whether such isolation is due to personal choices or to marginalization, (II) whether different demographic groups build their personal networks

differently over time. The study adopts the proposal that an online network evolves as a result of:

(i) *structural factors of the network e.g. reciprocity and centralization, and*

(ii) *demographic characteristics of participants such as education, gender, and personality*

The first research question is as follows:

RQ1: How do structural and demographic factors contribute towards the formation of online social networks in organizations?

Several social network theories e.g. social capital and social comparison have been proposed to explain how social networks form. Social capital theory posits that people form relationships to gain access to resources (Coleman 1988; Lin 1999). Social comparison theory, on the other hand, posits that people form relationships on the basis of similarity (Festinger 1954; McPherson, Smith-Lovin, and Cook 2001); hence, men are more likely to seek advice from other men, women are more likely to be friends with other women and so on. These theories were primarily generated through examination of offline social networks i.e. advice, friendship, club membership etc. (Cross, Borgatti, and Parker 2001; Currarini, Jackson, and Pin 2009; McPherson, Smith-Lovin, and Cook 2001). However, online networks might be different from offline networks for a variety of reasons.

First, greater effort is required to establish and maintain relationships in offline networks compared to online networks; typically, one needs to click a button to signal initiation of the relationship or acceptance of a relationship request in online platforms (Kane et al. 2012). Second, electronic tools are radically altering the structure of social

interactions; hence, geographic and social status boundaries are being destroyed, such that teams separated by oceans can successfully collaborate on complex software development projects, and private citizens can directly communicate with their national leaders online (Chiu, Hsu, and Wang 2006; Cummings, Butler, and Kraut 2002). Thus, in online networks, geographic and social boundaries may be less important than in offline networks. Third, online networks exhibit varying degrees of visibility to network members i.e. it typically requires some effort to determine who is related to whom within an online network whereas in offline networks relationships are more readily apparent to others. For these reasons, the structural characteristics of online networks might differ from those of offline networks. Our second research question:

RQ2: How do online social networks differ from offline social networks in organizations?

Online networks might also interact with offline networks. In organizations, each employee is likely to be simultaneously part of multiple social networks – both offline and online e.g. email, friendship, advice etc. (Agarwal, Gupta, and Kraut 2008; Oinas-Kukkonen, Lyytinen, and Yoo 2010). People's centralities in one network might be influenced by their centralities in another network; for example, influential people in the reporting network may also be influential in the communication network, because they are likely to control the flow of information (Burt 2002). While some studies have explored how offline networks such as advice and friendships networks correlate (e.g. Krackhardt 1990), an open question remains on whether online networks echo, complement, or substitute offline networks in organizations (Cummings, Butler, and Kraut 2002; Kane et al. 2012). Hence, this study also examines the nature of the interplay among online and offline social networks. The third research question:

RQ3: How do online social networks interact with offline social networks in organizations?

Thus, this study has three main goals: (i) uncovering the factors affecting online network formation (ii) revealing the structural differences between online and offline networks and (iii) examining the interplay among online and offline networks in organizations.

Answering our research questions contributes towards our understanding of online networks in particular and social networks more broadly. By examining the evolution of an email social network of a technology company over a two-year period, this study makes several contributions. First, we show that online network formation is affected by the personal characteristics of network members i.e. personality and education. Second, we show that online network formation is also affected by network members' positions in advice, brainstorming, and working networks. Third, we show that online network formation is also affected by broader network characteristics of reciprocity and centralization tendencies. The study makes further contributions in revealing the differences between online and offline networks in organizations. Methodologically, we outline a method of constructing an online network using email communication logs, and we also show how to accomplish this while protecting employee privacy.

In Chapter Two which comes next, we outline a brief history of social network research, to highlight some of its most prominent implications. We also outline the reasons why online social networks provide new opportunities for the advancement of social network research. Chapter Three theoretically justifies a set of hypotheses for

online social network formation and evolution, as well as for the differences between online and offline networks. Chapter Four outlines the methods of data collection and analyses. Chapter Five describes the results. The penultimate section, Chapter Six, discusses the results at length, offers some theoretical and practical implications, and precedes Chapter Seven which concludes with a summary of the study.

CHAPTER TWO

Background Research

In this chapter, we describe the fundamental concepts undergirding social network research. We begin with a broad overview of social network foundational concepts and network concepts. We then provide a brief recounting of social network history, before highlighting how the social network perspective has been applied in research in organizational studies and information systems. We conclude the section by showing the utility of online networks in answering hitherto difficult questions in social network research.

Foundational Concepts

A social network is defined as a set of entities with relationships among them (Wasserman and Faust 1994). Entities are people or groupings of people e.g. teams, organizations, and nations. Social networks are depicted by drawing links between any pair of people (or groupings of people) with relationships among them; these links are called ties or edges. Ties can be directed or undirected; on Facebook, for example, for a tie to exist between two people, one of them sends a friendship request and another accepts the request; hence, ties on Facebook are undirected. On Twitter, on the other hand, where person A follows person B, a tie would be drawn from A pointing towards B. Bidirectional ties are also possible e.g. on Twitter, if both A and B follow each other, a bidirectional tie would be drawn between them. Ties vary in their strengths, for example in email networks, if A tends to communicate with both B and C, but communicates with

B more often than with C, then the line from A to B will be thicker than the line from A to C. Figure 2.1 illustrates these concepts. By drawing the social network, one may discover paths linking multiple people that are not directly connected to each other.

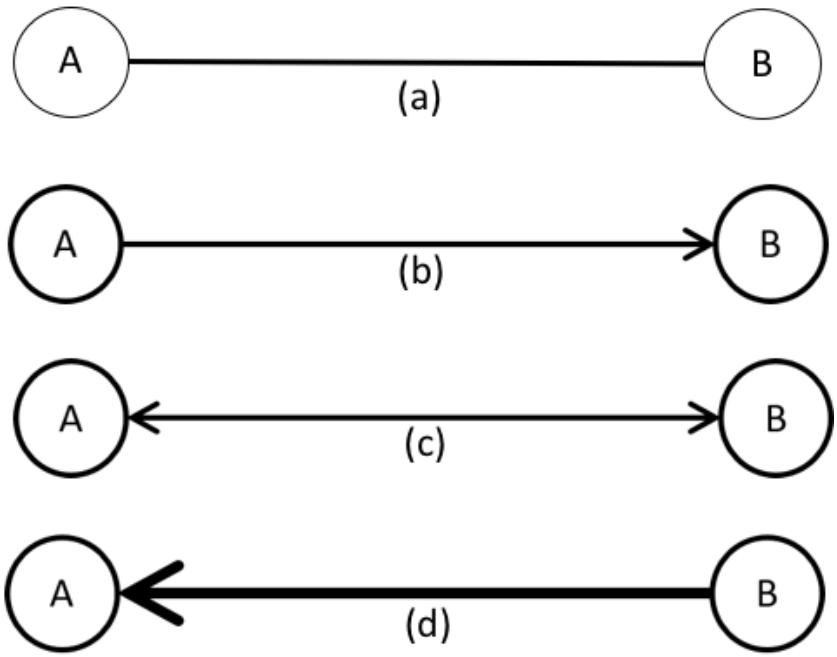

Figure 2.1. Social network notation. (a) Undirected tie (b) Unidirectional tie (c) Bidirectional tie (d) Strong directed tie

Network Characteristics

A variety of measures are used to describe the structural characteristics of networks, which allows for comparison of multiple networks. For example, the networks of different groups might be compared by their sizes i.e. the number of network members. Particular network characteristics are associated with various outcomes, and below we explore a variety of these measures.

Network Size

Social network size has been evaluated at the individual and network level. Size at the individual level refers to the number of ties in an individual's personal network. Studies from evolutionary biology indicate that personal network size is constrained by the sizes of certain parts of the brain, particularly the neocortex and amygdala, such that the greater the sizes of these brain sections, the larger an individual's network size (Bickart et al. 2011; Dunbar 1992). The larger the neocortex of an individual, the greater the information-processing ability of the individual and the greater the number of relationships the individual can maintain (Dunbar 1992). The mean number of ties maintained by humans in offline networks is estimated to be 150, beyond that the information processing required to maintain ties becomes too expensive for the neocortex and amygdala. Personal network size decreases with age; as people get older, they lose relationships because their contacts die, and because their capacities to maintain relationships decreases with age (Bickart et al. 2011; Van Tilburg 1998). How might this situation differ in online networks? Use of electronic communication tools may increase the sizes of personal networks, because rather than being carried out by the neocortex or amygdala, the information processing required to maintain relationships can be delegated to these tools.

Size can also be measured at the network level. In this context, size refers to the number of network participants. Within a single organization, departments are likely to vary in size, as are groups or teams tasked with carrying out certain projects. Team size predicts team performance (Reagans and Zuckerman 2001). Employing the social network perspective shows that larger teams leverage their greater social capital relative

to smaller teams to maximize their performances. In scientific collaboration networks, larger research teams are more productive and also produce higher impact research than smaller teams (R. B. Freeman and Huang 2015; Gallivan and Ahuja 2015). Electronic tools can bring together people separated by long geographic and social distances; thus, they can increase team sizes and increase collaboration leading to better performance and increased research impact.

Network Density

Another way in which networks differ is in their densities. Density is calculated by counting the number of ties and dividing it by the potential number of ties (Wasserman and Faust 1994). The potential number of ties is dependent on the number of nodes in the network i.e. its size. A saturated network i.e. where every node is connected to every other node, has the following maximum possible number of ties in an undirected network:

$$\text{Number of potential ties} = \frac{n(n-1)}{2}$$

where n is the size of the network.

Density is a measure of network cohesion, because it measures the degree of cooperation among nodes within a network (van Beek et al. 2011). Density figures can be misleading when comparing networks of different sizes. As the size of a network grows, the maximum potential number of ties in the network grows exponentially (see equation above). For large networks, the densities are likely to be very low; hence, networks with large differences in size are inappropriate to compare using raw density figures (Wasserman and Faust 1994). For small networks of comparable size, density can be

informative; for example, density indicates the extent to which a group is cohesive, with higher density showing more cohesive networks (Blau 1977). Density is also easily interpretable when comparing different networks occupied by the same set of people e.g. communication vs. advice networks within a single organization. In this regard, differing densities show the relative difficulty of initiating and maintaining ties across networks. Our study examines how the densities of offline and online networks differ.

Reciprocity

For directed social networks, reciprocity assesses the extent to which, for any two individuals A and B, if A sends a tie to B implies that B sends a tie to A (Hanneman and Riddle 2005). In offline networks reciprocity is measured using two methods: 1) asking network members their feelings about every other member, and 2) asking network members direct questions about reciprocity i.e. their perceptions of how reciprocal the relationships between them and every other network member are. The measurement of reciprocity in offline networks heavily bases on perceptions; hence, it is not immune to limitations of self-recall and social desirability bias. In online networks these limitations are easily obviated, but this advantage comes at the cost of more nuanced assumptions about measuring reciprocity.

Measuring reciprocity in online networks requires some nuance. On platforms such as Twitter and Instagram, one has to make a conscious effort to follow another and accepting a follow request does not imply following the requester; hence, measuring reciprocity in this regard is not difficult. But on platforms such as Facebook and LinkedIn, a responder accepting a request means that the requester and responder automatically become friend and connection on the two platforms respectively. Hence,

the directed networks on Facebook and LinkedIn offer no information about reciprocity because a tie in one direction implies a tie in the other direction in all cases. To gain a meaningful picture of reciprocity between any two members in such networks, it is more instructive to look at patterns of interactions within these platforms, for example by examining the number of comments, likes, and message exchanges between the network members. For example, if person A tends to send many more messages to person B than B sends to A on Facebook, the relationship may be deemed not reciprocal, but if A and B tend to send roughly equal numbers of messages to each other, the relationship may be deemed reciprocal. Thus, reciprocity is a concept that requires nuanced treatment in online network settings, further underscoring the differences between online and offline networks and by extension, justifying the utility of our study.

Reciprocity is important to study because it is a measure of the stability and institutionalization – the respect afforded to social relations and organizational norms – of the network (Hanneman and Riddle 2005). Higher reciprocity implies higher network stability and institutionalization. Networks with low reciprocity have majority asymmetric ties; such ties are not stable and are bound to dissolve with time. Reciprocity can have important financial implications; increases in director compensation over time are partly explained by reciprocity between directors in interlocking company boards of directors (Boivie, Bednar, and Barker 2015), and employees apply effort in proportion to how hard their co-workers work; in other words, they respond not only to wages but also in reciprocation to their perception of co-worker efforts (Gächter, Nosenzo, and Sefton 2012). It is unclear whether levels of reciprocity should be higher in online or offline

networks, given the reduced levels of effort required to reciprocate relationships in these networks.

Centrality/Centralization

Individuals' positions within social networks afford them some measure of influence over the way information and other resources flow within the network. In social network terms, network position is called centrality and is a measure of network influence (L. C. Freeman 1978); hence, an individual on the outskirts of the network yields less influence than one within the center of the network.

Various types of centrality exist (for more detailed coverage, see Koschützki et al. 2005); in this study, we will explore the four most influential types. The first type is called *degree*. A node's degree is its number of direct ties i.e. the number of nodes that lie adjacent to it in the network (Nieminen 1974). In directed networks, a further distinction is made between a node's *outdegree* or *gregariousness* – its number of outgoing ties – and *indegree* or *popularity* – its number of incoming ties (Godde et al. 2013; Wasserman and Faust 1994). Degree is indicative of activity; nodes with high degree are actively involved in network activity. Such nodes act as channels of information and the higher the degree of a node, the more focal the node as a point of communication (L. C. Freeman 1978).

A second type of centrality is called *betweenness*. Betweenness is the extent to which a node lies in between other nodes within a network (Hanneman and Riddle 2005). In this regard, a node is central in the network to the extent that it lies on the shortest paths between pairs of all other nodes in the network (L. C. Freeman 1977). The node with the highest betweenness centrality is important to the network because its removal

penalizes every other node's average path length to other nodes the highest. Path length refers to the number of ties that lies between two nodes (McPherson, Smith-Lovin, and Cook 2001). Nodes with high betweenness draw their influence from their abilities to withhold information or distort it during transmission between the nodes that they link together (Bavelas 1948). Additionally, such nodes can also coordinate activities among nodes in disparate areas of the network, owing to their ability to control the flow of information (Cohn and Marriott 1958).

Closeness is another centrality type. Closeness is the extent to which a node is near to other nodes in network distance terms (Okamoto, Chen, and Li 2008). It is calculated by taking the inverse of the node's average path length to other nodes. Hence, the node with the shortest average path length will have the highest closeness centrality. Nodes with high closeness centrality can spread information (or whatever flows within the network) to other nodes with higher efficiency than nodes with low closeness centrality (Borgatti 1995).

A fourth type of centrality is *eigenvector* centrality. Eigenvector centrality is the extent to which a node is connected to influential nodes within the network (Bonacich 2007). The intuition behind this measure is that nodes are not likely to exercise equal influence within a network, hence it might be more instructive to weigh a node's centrality according to the centralities of the nodes that it connects to (Bonacich 1972). A node's eigenvector centrality is proportional to the sum of the centralities of each of its adjacent nodes; in this regard, a node's eigenvector centrality is calculated through a recursive mechanism (Borgatti 1995). Unlike degree centrality which only considers the

number of direct ties, eigenvector centrality also considers the centralities of indirect nodes.

Measurement of centrality is likely to be more accurate in online networks than offline networks, which mainly rely on survey methods for collection of relational data. Some survey-based social network studies, because of practical concerns, have asked people to nominate up to a specific number of people for the purposes of mapping out the social networks at play within the network (e.g. Ibarra 1992; Kardos et al. 2017). In that case, most people will have an outdegree of whatever the artificial limit is, which distorts the true degree centrality picture. In such cases, eigenvector centrality suffers from the same limitations, because it will not differ from degree centrality (Bonacich 2007). With online networks, one may just consult the digital trace data e.g. email archives and use the metadata to reconstruct the networks; this may enable true and more informative centralities to be calculated.

Centrality varies widely among nodes. Centrality has typically been employed as an antecedent of various outcomes i.e. innovation, work performance, employee well-being etc. (Agneessens and Wittek 2008; Burr 1992; Gibbons 2004; Ibarra 1992); however, in recent decades the distribution of centrality in networks has attracted more attention in the literature, with various studies demonstrating that a long-tailed degree distribution is a feature of complex networks including social networks (Albert, Jeong, and Barabási 1999; Barabási and Bonabeau 2003; Newman 2001a). The long-tailed degree distribution means that the majority of nodes in a network have low degrees, and a few nodes will have high degrees. One can compare the degree distributions of two networks by examining the exponent of each network's degree distribution – the higher

the exponent the more dominant the high degree nodes in determining the average degree per node (Newman 2001a). To further our understanding of what is different about online networks (Kane et al. 2012), our study compares the degree distributions of online and offline networks.

Homophily

A fifth characteristic that differentiates networks is homophily – or the extent to which individuals form ties with others similar to themselves (Currarini, Jackson, and Pin 2009). Examples of such characteristics include gender, race, religion, age etc. The informational theory of homophily suggests that people prefer forming ties with similar others because it is easier to communicate with them than with different others (Egorov, Polborn, and Welcome 2010). The self-categorization principle also explains homophily; the principle states that individuals place themselves in different categories based on certain characteristics i.e. gender, race, intelligence etc. (Reynolds 1987). If they perceive others to be in a similar category, they subsequently prefer interacting with these similarly categorized people than with those in different categories. Thus, categorization becomes the basis for homophily.

Homophily has been demonstrated in a variety of networks including friendship, advice, confiding, and marriage networks (Currarini, Jackson, and Pin 2009; Ibarra 1992; Kalmijn 1998; Lazarsfeld, Merton, and others 1954; McPherson, Smith-Lovin, and Cook 2001). However, for these networks, the physical cues necessary for categorizing others were present. When seeking advice or marriage for example, the physical attributes (age, gender, or race) of a possible partner are readily observable. For online networks, such information might not be available and when it is availed, through the use of profile

pictures for example, the information might not be accurate. Hence, it is unclear whether homophily should persist in online networks or not. Our study further sheds light on whether homophily exists in online networks. Table 2.1 shows the definitions of network characteristics.

Table 2.1. Definitions of network characteristics

Term	*Description*
Size	At the individual level, a node's number of ties (Dunbar 1992). At the network level, the number of nodes in a network
Density	The proportion of observed ties relative to the maximum number of ties (Wasserman and Faust 1994).
Reciprocity	The extent to which any node *i* with an outgoing tie to another node *j* has an incoming tie from that node *j* (Hanneman and Riddle 2005).
Centrality	The position of a network within a network; it is a measure of network influence (L. C. Freeman 1977)
Homophily	The extent to which similar nodes form ties compared to what is predicted by chance (McPherson, Smith-Lovin, and Cook 2001).

Having described the characteristics that distinguish different networks, we now recount a brief history of social network research. In particular, we highlight the importance of social network centrality and how it influences multiple individual and organizational outcomes of interest. Centrality within a network determines an individual's access to resources held by network members (Burt 2002; Lin 1999). Recounting this history shows that social network research has largely focused on the influence of individuals' structural positions on outcomes, but has little explored the antecedents of those positions.

A Brief History of Social Network Research

Early work on social networks examined relationships among small groups of people i.e. classrooms, workplaces, clans etc. (J. L. Moreno 1934; Simmel 1903). In this work, human relationships were much constrained by geographic and status boundaries, such that the importance of familial and community relationships trumped all others'; this is captured in Moreno's quote "the nearer two individuals are to each other in space, the more do they owe to each other their immediate attention and acceptance" (J. L. Moreno 1934, 58:xx). Work in sociology and anthropology further underscored how social networks exhibit segregation across racial, religious, and educational boundaries (Malinowski 1954, 2002; McPherson, Smith-Lovin, and Cook 2001). Hence, early work largely ignored how the smaller local networks aggregated to create the broader social networks that form societies.

Stanley Milgram, in his famous "small world" experiment, was one of the first to show that social networks link people together across much greater geographic distances than previously thought (Travers and Milgram 1969)[1]. He showed that in the US, a nation of hundreds of millions, any two individuals are separated by only six people when the social network of its population is considered. Further work linking local networks to larger networks showed that people with strong ties are more likely to share multiple mutual acquaintances than people sharing weak ties (M. Granovetter 1983). This variation in strength has implications for multiple outcomes because individuals are more likely to receive novel information through their weak ties than their strong ties – a

[1] In the experiment, several people from Midwestern US city of Wichita, Kansas were each asked to send a folder to a woman in the Northeastern US city of Boston, Massachusetts; results showed that on average, the folder exchanged hands five times before reaching the intended recipient.

phenomenon dubbed the “strength of weak ties”. This strength of weak ties is quite pervasive; it implicates knowledge transfer in organizations (Hansen 1999), mediation of information about opportunities for new jobs (Brown and Konrad 2001), and even information on crime opportunities in the underworld (Coles 2001; Patacchini and Zenou 2008). With electronic tools increasingly mediating human interactions, the collection of tie strength information will be less difficult, making possible the investigation of more complex questions.

Like the strength of weak ties, structural holes form another social network concept that implicates multiple outcomes. Structural holes are areas that link disparate parts in a network (Burt 2002). In network centrality terms, individuals occupying structural holes have high betweenness centrality (Burt 2002). The concept of structural holes arises out of the social capital theoretical tradition i.e. one’s connectedness within a social network can be considered an asset, and higher connectedness signifies greater social capital. Individuals and social groupings occupying structural holes play significant roles within the network, because they broker knowledge transfer within the network. As such, researchers occupying structural holes in scientific collaboration networks tend to become editors in influential journals (Gallivan and Ahuja 2015; Xu, Chau, and Tan 2014); organizations occupying structural holes within the U.S. health policy network enjoy greater influence than others (Fernandez and Gould 1994); and firms occupying structural holes in the network of joint ventures tend to be innovative, in that they produce more patents than other firms (Ahuja 2000). How individuals and social groupings come to occupy these structural holes is little examined, and this study answers this question by examining network formation.

Table 2.2 summarizes the impact of structural position on individual and organizational outcomes.

Table 2.2. Structural position and its impacts.

Structural position	*Centrality type*	*Outcomes*
Structural hole (Burt 2002)	Betweenness	Organizational innovation (Ahuja 2000), organizational influence (Fernandez and Gould 1994), scholarly influence (Xu, Chau, and Tan 2014)
Diversity of ties (M. S. Granovetter 1973)	Degree	Knowledge transfer in organizations (Hansen 1999), access to information on job opportunities (Brown and Konrad 2001), access to information on crime opportunities (Coles 2001).

Presently, the social network perspective is pervasive. In the medical field, it has been shown that personal network characteristics have a genetic explanation. One intriguing finding is that an individual's popularity within a network can be partly explained by the individual's genetic composition and is heritable (Fowler, Dawes, and Christakis 2009). In economics it has been shown through social network analysis that the level of effort exerted by employees not only depends on their wages, but on other employees' wages as well (Boivie, Bednar, and Barker 2015; Gächter, Nosenzo, and Sefton 2012). Social networks have also been implicated in the diffusion of happiness, smoking, and obesity across societies (Christakis and Fowler 2007, 2008; Fowler and Christakis 2008). The social network perspective is thus powerful and implicates multiple important outcomes.

This section presented a general overview of the history of social network research, and explored the influential concepts of structural holes and strength of weak ties and how they implicate multiple outcomes e.g. work performance and innovation

respectively. We highlighted some possibilities enabled by online networks, particularly the easier calculation of tie strength and investigation of antecedents to structural holes. Our study lies at the intersection of organizational studies and information systems (IS). Thus, in the next two subsections, we provide more specific descriptions of how the social network perspective has been applied in organization studies and IS research.

Social Networks in Organizations

Because organizations are groupings of people with relatively stable patterns of interactions across time (Tichy, Tushman, and Fombrun 1979), the social network perspective has long been influential in organizational studies. The perspective has been used to examine organizational change, communication, trust, and unethical behavior (Brass, Butterfield, and Skaggs 1998; Chow and Chan 2008; Tichy, Tushman, and Fombrun 1979). The utility of studying social networks in organizations is in linking individual characteristics to broader organizational behavior. An example is the question why organizations engage in unethical behavior – one school of thought attributes this behavior towards the personal characteristics of organization members i.e. this behavior is the result of a few "bad apples"; a second school of thought attributes such behavior towards organizational norms, codes of conduct, reward systems etc. that influence otherwise ethical individuals i.e. "bad barrels" (Brass, Butterfield, and Skaggs 1998, 14). The social network perspective adds the influence of relationships among organization members as an explanation of why otherwise ethical organization members might engage in such conduct. Thus, by examining who influences whom within the organization, a more complete picture of organization behavior can be revealed.

Most social network research in organizations examines the outcomes associated with one's social network position and diversity of ties. Position refers to various centrality measures, the most prominent of which are structural holes and popularity (a node's number of incoming ties). Individuals occupying structural holes demonstrate higher innovation than individuals that do not, primarily because these individuals link together disparate parts of the network and have access to different ideas that they can combine to produce innovative ideas (Carnabuci and Diószegi 2015). Promotions, higher compensation, and greater job performance are benefits that also accrue from occupying structural holes (Burt 2004). Popularity in advice and communication networks is positively associated with job performance, job security, and organization commitment (Zhang and Venkatesh 2013). Individuals with greater tie diversity i.e. a mix of strong and weak ties are also more innovative (M. Granovetter 1983).

To summarize, individuals' structural positions within social networks can afford them tangible advantages. However, how structural position emerges is still understudied (Borgatti et al. 2009; Johnson, Faraj, and Kudaravalli 2014). By studying the history of social networks and the application of the social network perspective in organization studies, we observe that most social network research is focused on outcomes of social network position. Hence, occupying structural holes and maintaining a diversity of ties affords an individual many advantages. Yet the factors that give rise to social networks remain understudied. An improved understanding of such factors helps reveal whether certain groups enjoy advantages in occupying these positions or not, which might be useful for organizations as they seek to understand these consequential structures. To that

end, we leverage technology in the form of communication logs to uncover how social networks form.

Social Networks in IS Research

An individual's centrality in online social networks implicates multiple variables of interest. In organizations, an individual's number of offline and online ties positively influence job performance (Zhang and Venkatesh 2013). The introduction of social media into an organization can alter network structure in ways that may benefit some individuals and disadvantage others (Wu 2013). Centrality within an organization's electronic bookmarking social network[2] is positively associated with innovation at the individual level (Gray, Parise, and Iyer 2011). Highly central IT systems achieve better efficiency and quality of care outcomes in health organizations than lowly central systems (Kane and Alavi 2008). Groups whose most proficient members in specific IT systems are highly central perform better than groups whose most proficient members have low centrality (Kane and Borgatti 2011). Individuals that are highly central in email networks are at higher risk of infecting other computers with viruses than lowly central individuals (Guo, Cheng, and Kelley 2015).

At the firm organizational level of analysis, firm centrality in networks of organizations influences several outcomes. Firms highly engaged in value co-creation activities through strategic alliances with other firms can enhance their performances (Gnyawali, Fan, and Penner 2010). In open source service networks, in moving products

[2] A bookmarking system allows individuals to create electronic links to digital resources and assign metadata to these links. The ensuing social network embeds the relationships created when other individuals within the organization access these bookmarks.

from the peer production phase to productization, access to and transfer of key strategic resources among member firms must be encouraged through an emphasis on collaboration and sharing (Feller et al. 2008). In crisis response situations, a variety of information response network prototypes are possible based on the amount of information intensity and reach of network density; the star prototype emerges when both information intensity and network density are high, the information pyramid is characterized by high information intensity and low network density, the information forest is characterized by both low information intensity and network density, and the information blackout is characterized by low information intensity and high network density (Pan, Pan, and Leidner 2012).

In online social networking sites, experiments show that the stronger an existing tie between two people on Facebook, the more likely they are to cooperate in economic settings (Bapna et al. 2011). On YouTube, social interactions such as subscriber and friendship relationships determine the success of videos and the magnitude of that success (Susarla et al. 2012). When Oprah Winfrey endorses a book for her book club, not only does the book enjoy higher sales, books that are recommended along with it on Amazon also enjoy higher sales (Carmi, Oestreicher-Singer, and Sundararajan 2012). Hence online social networks have important implications for multiple social and economic outcomes.

Although much is known about these effects of social networks, little empirical research exists on the mechanisms that lead to the formation and evolution of organizational social networks. Most notably, network dynamics over time have been understudied (Borgatti et al. 2009). Nevertheless, a few studies have explored online

network formation and evolution. In a study of 28 online communities, Johnson et al. (2014) found preferential attachment to be an inadequate explanatory mechanism for online community social network formation; the mechanisms of least efforts (where more prepared individuals are more likely to contribute than everyone else), direct reciprocity, and indirect reciprocity (where individuals contribute back towards a group rather than a specific individual) also contribute towards network formation. Yan et al. (2015) examined network formation in an electronic network of medical patients. They found that patients with similar firsthand disease experience were more likely to establish relationships than patients with dissimilar experiences. Further, patients' cognitive abilities, such as the information load and range of social ties they can manage, were found to limit network growth. Our review shows that the formation and evolution of online social networks in organizations have not received attention in the IS literature, hence the results of our study are useful in revealing the factors fueling the dynamics of such networks within organizations. Table 2.3[3] shows a selection of the social network research carried out in IS journals over the last decade.

Of the 40 studies constituting the sample of articles described in Table 2.3, the bulk (60%) examined impacts of social networks on outcome variables, 20% examined network formation, and a further 20% were conceptual and/or review articles (Figure 2-2). This distribution of articles shows that relative to studies on network impact, network formation is much less studied in IS research.

Our study differs from prior studies on network formation in two important ways.

[3] This selection of articles is obtained from the Basket of 8 journals (*MISQ, ISR, JMIS, JAIS, JSIS, EJIS, JIT*, and *ISJ*), *Management Science*, *Organization Science*, *Social Networks* and others.

Table 2.3. Recent IS research on social networks

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Bapna, Gupta, Rice, and Sundararajan 2011)	Experiment/Individual	Strength of social ties	Trust, reciprocity	There is significant correlation between strength of tie and trust and reciprocity. Cooperation is likely when a tie exists between two people.
(Carmi, Oestreicher-Singer, and Sundararajan 2012)	Quantitative - network construction using Amazon hyperlinks, SNA/Product	Local clustering, indegree, network proximity	Spillover	Sales of products up to 3 units of geodesic distance away from a product experiencing an exogenous shock are positively affected by the shock.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Dissanayake, Zhang, and Gu 2015)	SNA/Team	Social capital, intellectual capital	Team performance	On an online crowdsourcing platform, competition moderates the impact of social-intellectual capital alignment on team performance. In high-intensity competition contexts, such alignment has negative impact, while for low-intensity competition it has positive impact.
(Dou, Niculescu, and Wu 2013)	Econometric model/Individual	Seeding.	Adoption, willingness to purchase.	Under complete information, disutility can result from seeding.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Fang, Hu, Li, & Tsai, 2013)	Bayesian prediction/Individual	Social influence, structural equivalence, entity similarity, confounding factors	Adoption probability.	Beyond social influence, hidden confounding factors also influence adoption decisions.
(Faraj and Johnson 2011)	Exponential random graph modeling (ERGM), Mixed ANCOVA/Online community	Direct reciprocity, indirect reciprocity	Network	Online communities demonstrate direct and indirect reciprocity tendencies, but not preferential attachment.
(Faraj, Kudaravalli, and Wasko 2015)	Survey, content analysis, SNA/Individual	Sociability, knowledge contribution, structural social capital	Leadership	Structural social capital, knowledge contribution, and sociability positively influence leadership in online communities.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Feller, Finnegan, Fitzgerald, and Hayes 2008)	LISREL, content analysis/Organization	Coordinating exchanges, safeguarding exchanges	Access to and transfer of strategic resources.	Coordinating exchanges and safeguarding exchanges are positively associated with access to and transfer of strategic resources in open source service networks. Informal social mechanisms should be employed in the design of business networks in order to deter unhelpful behavior from network members.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Gnyawali, Fan, and Penner 2010)	OLS regression/Organization	Value co-creation actions (co-development and relational capability), Repertoire of actions (complexity, volume)	Firm performance	Co-development actions are positively related with firm performance.
(Godinho de Matos, Ferreira, and Krackhardt 2014)	Agent-based modeling/Individuals	Peer influence	Diffusion of adoption	Peer influence positively affects adoption. However, offering free products to influential nodes may not be profitable.
(Gnyawali, Fan, and Penner 2010)	SNA/Individual	Outdegree, network effective size (the extent to which ego's network is likely to yield non-redundant information),	Innovation	Higher diversity in sought information is positively associated with individual innovation.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Guo, Cheng, and Kelley 2015)	SNA/Individual	Individual betweenness, group size	Malware propagation	Betweenness centrality and membership in large groups is associated with higher risk of infecting other computers.
(Hahn, Moon, and Zhang 2008)	Logistic regression/Individual	Prior collaboration, perceived status of project members	Project joining	Past collaboration positively influences the decision to join a team.
(Hinz and Spann 2008)	Lab experiment/Individual	Information diffusion, betweenness centrality	Bidding behavior	Betweenness centrality has impact on bidding behavior.
(Howison, Wiggins, and Crowston 2011)	Research commentary			Digital traces may provide invalid data if data is taken at face value without considering the context of use.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Johnson, Faraj, and Kudaravalli 2014)	Agent-based modeling/Individual/Dyadic	Preferential attachment	Network	Preferential attachment is inadequate to explain the power law distribution in online communities.
(Kane and Borgatti 2011)	SNA/Group	Centrality-IS performance alignment	Performance	Centrality-resource alignment is more useful than simply averaging individual resources. Highly central nodes may affect group behavior more intensely.
(Kane, Alavi, Labianca, and Borgatti 2012)	Literature review			Online and offline networks interact in that ties in these networks may complement, echo, or substitute for the other.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Kane and Alavi 2008)	Multimodal SNA/Group	Average user-system tie strength, IS centrality	Efficiency, quality of care	IS centrality is significantly related to both efficiency and quality outcomes at the group level
(Kane and Ransbotham 2016)	Affiliation SNA/Individual article	Centrality	Article quality	The position of an article in the affiliation network is associated with the quality of the article
(Kudaravalli and Faraj 2008)	OLS and logistic regression/Group	Initiating dialogue, sustaining dialogue	Effectiveness of collaboration	Initiating dialogue and sustaining dialogue impact the success of discussion threads.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Leonardi 2013)	QAP, ethnography/Group	Technology feature use	Social network change	Informal organization networks only change due to technological change when the affordances of introduced technologies are jointly realized.
(Leonardi 2015)	OLS regression, t-test/Individual	Exposure to Content of Others' Messages, Exposure to Indicators of Others' Communication Partners	Meta-knowledge - accuracy of who knows what, accuracy of who knows whom	Use of enterprise social networks increase accuracy of meta-knowledge i.e. the knowledge of "who knows what" and "who knows whom".

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Levina and Arriaga 2014)	Theory building			In online fields, users drawn to particular communities based on common interests are also differentiated according to status.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Oinas-Kukkonen, Lyytinen, and Yoo, 2010)	Research commentary			The IS research drawing on social networks can be divided into the following streams: 1) network awareness at both individual and organizational levels, 2) uses of social network analysis related to IS use, and 3) conceptual and technological change in the fast-evolving platforms to manage social networks.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Pan, Pan, and Leidner, 2012)	Case study/Organization	Information intensity, network density, direction of info flow and role of information response organization	Type of information network	Four types of information flow networks emerge based on information intensity, network density, direction of information flow and role of information response organization: star, pyramid, forest, and blackout.
(Park, Huh, Oh, and Han 2012)	Simulation, inference model/Individual	Relational data	Validity of self-reported demographic data	Homophily based inference model outperforms competing models in identifying discrepancies in self-reported demographic data.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Polites and Watson 2009)	SNA/Journal	Citation relationships	Journal prestige, centrality	The IS discipline is a net receiver, rather than a net provider, of information from other disciplines.
(Putzke, Fischbach, Schoder, and Gloor 2010)	ERGM, SIENA/Individual	Structural and demographic variables	Network	Transitivity, reciprocity, and homophily influence network evolution. Structural embeddedness in the interaction network does not influence player performance.
(Qiu, Rui, and Whinston 2014)	Randomized experiment/Individual	Degree, peer effects	Performance, outside information acquisition	Participants with higher degree are likely to not seek information from outside sources.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Ransbotham and Kane 2011)	Cox proportional hazard/Individual	Membership turnover	Effective collaboration - knowledge creation and knowledge retention	Knowledge creation and knowledge retention constitute separate stages of collaboration in online communities. Membership turnover is desirable to an extent because it encourages generation of novel information and abilities
(Ridings and Wasko 2010)	Mixed methods/Online community	Communication volume, message length	Member retention, (new) member attraction	Over time, increased efforts by members leads to greater retention, but new member attraction is diminished.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Robert Jr, Dennis, and Ahuja 2008)	Lab experiment/Individual	Social capital (structural, relational, cognitive)	Knowledge integration and performance	Relational social capital positively influences knowledge integration equally in online and offline interactions. Structural social capital and cognitive social capital had greater impact on knowledge integration in online compared to offline interactions.
(Sarker, Ahuja, Sarker, and Kirkeby 2011)	SNA/Team	Communication, trust	Performance	Trust mediates the effect of communication on performance.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Schaefer, Light, Fabes, Hanish, and Martin 2010)	SIENA/Individual, dyadic	Reciprocity, popularity, triadic closure	Friendship social network	Reciprocity and outdegree density were related to network evolution and formation.
(Shi, Rui, and Whinston 2013)	Conditional maximum likelihood estimation/Individual	Tie strength	Content sharing (retweet)	An individual's weak connections are more likely to share the individual's content with their neighbors than the individual's weak ties.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Susarla, Oh, and Tan, 2012)	Bass model/Video	Centrality in subscriber network	Diffusion of video post	Social interactions determine the success and magnitude of posted content on YouTube. Indegree centrality is positively associated with diffusion of video posts.
(Wu 2013)	OLS/Individual	Structural diversity.	Productivity, job security	Introduction of social media in an organization can alter centrality positions within a network. An individual's structural diversity positively influences his/her productivity and job security.

Citation	*Methodology/Level of Analysis*	*Independent Variable*	*Dependent Variable*	*Key Conclusion*
(Yan, Peng, and Tan 2015)	ERGM/Individual	Individual control, cognitive capabilities, disease experience-based similarity	The propensity to establish/maintain a social relation	The more embedded a patient is in the network of patients, the less likely that she will establish new social relations.
(Zhang and Venkatesh 2013)	Survey/Individual	offline ties, online ties	Employee performance	Offline and online ties complement each other to positively affect job performance.

First, prior studies generally focused on the influence of structural characteristics on formation (e.g. Faraj and Johnson 2011; Ridings and Wasko 2010); in contrast, our study examines the influences of structural factors, demographic factors, and positions in offline networks on formation of online networks. Second, prior studies mainly examined network formation in online communities (e.g. Faraj and Johnson 2011; Faraj, Kudaravalli, and Wasko 2015; Yan, Peng, and Tan 2015); our study examines online network formation in an organization. In the next section, we outline how online networks can help ease processes of data collection that are necessary for studying network formation before explaining why studying network formation is important.

The Utility of Online Social Networks in Social Network Research

Existing research already outlines multiple reasons why the social network perspective is important. From the effect of centrality on outcomes such as job performance, job security, turnover intention etc., to the spread of influence in purchase decisions, technology adoption, happiness, smoking etc., social networks offer intriguing explanations for phenomena. Even then, how this pervasive structure materializes attracts relatively little attention in the literature. Social networks by nature change: relationships among people form, dissolve, strengthen and weaken, but this process is difficult to observe (Borgatti et al. 2009). One would need to repeatedly survey members of an organization to understand how these relationships change; even then some changes might escape the detection of the researcher, for example if a relationship frays and restores in between two data collection episodes. Online networks potentially address this shortcoming of traditional network data collection.

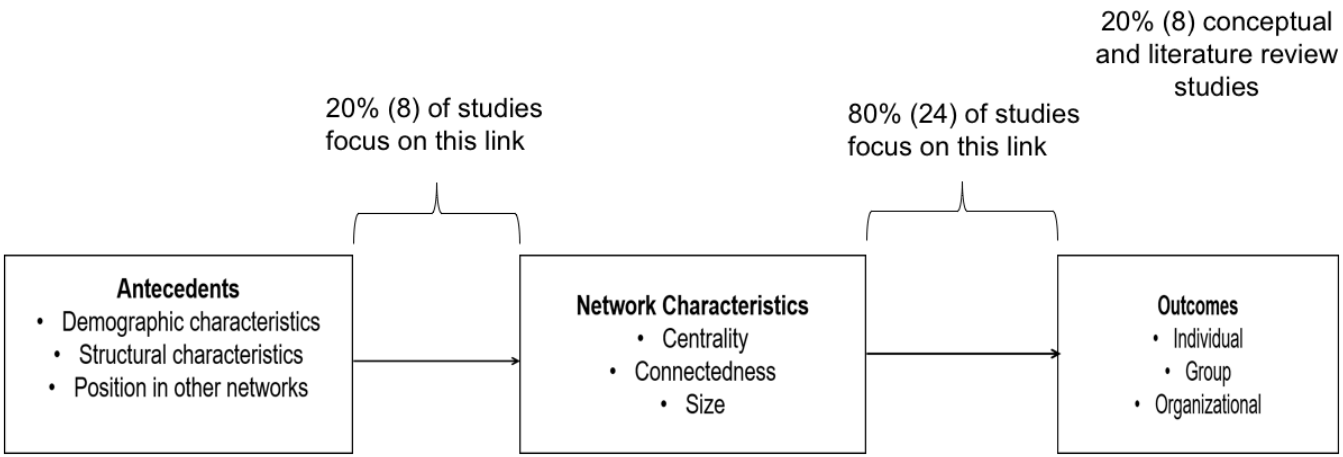

Figure 2.2. Distribution of studies according to research focus

Beyond mediating the distribution of valuable information within the organization, electronic tools enable the investigation of some complex social network phenomena. There are at least three possibilities enabled by the detailed records of electronic interactions. First, communication logs capture almost all electronic communication interactions between any pair of organization members, enabling more accurate depictions of the true nature of communication relationships within the organization (Oinas-Kukkonen, Lyytinen, and Yoo 2010; Putzke et al. 2010). Second, these interactions are timestamped, meaning that one can examine network dynamics without having to repeatedly survey organization members on their ties with other members. Third, the strengths of ties among organization members can also be captured by aggregating the number of interactions between pairs of organization members, which can be useful in understanding the dynamics of tie strengths. By capturing network dynamics, it becomes possible to answer multiple questions such as how social networks change in relatively normal times, how they change in times of crises, and most importantly what factors lead to such changes (Diesner, Frantz, and Carley 2005). From past work on traditional social network research, we understand that the following tendencies are consistent features of social networks.

First, social networks in general exhibit reciprocity, or high probability that two people participating in a relationship mutually agree that the relationship exist (Bapna et al. 2011; Gächter, Nosenzo, and Sefton 2012; Wasserman and Faust 1994). Second, social networks demonstrate preferential attachment i.e. ties accrue disproportionately to popular individuals within the network (Albert, Jeong, and Barabási 1999; Barabási and Bonabeau 2003). Third, they also demonstrate homophily, or a tendency for nodes to

form more ties with similar others than is expected of chance (Festinger 1954; Marsden 1988; McPherson, Smith-Lovin, and Cook 2001). While these tendencies are demonstrated to exist in traditional offline networks, it is unclear whether they should persist in online networks, or whether they are forces that actually shape the formation of networks. Longitudinal analyses of networks help answer some of these questions, and the use of IS tools greatly simplifies the task.

Table 2.4 shows a condensed summary of the social network literature, and some of the gaps existing thereof. Our study addresses two gaps: (i) whether existing social network theories generalize to online networks, or whether new theories should be created to understand them and (ii) identifying the structural and demographic factors that influence online network formation, and whether these factors differ for offline networks.

These gaps are important because network formation is not random (Barabási and Bonabeau 2003; Johnson, Faraj, and Kudaravalli 2014; Newman 2001b); rather, individual and contextual factors exert some control over how networks form. As such, if network formation arises out of certain human biases i.e. towards gender and/or ethnic preferences, some individuals might be systematically excluded from instrumental organization networks, which ultimately prevents them from getting access to new project and promotion opportunities (Agneessens and Wittek 2008; Lin 1999). Our study examines four networks within the organization, one online and three offline, to understand the differences between the two classes of networks and how they may interact with each other. The findings from this study address the important gaps identified from the literature, and Table 2.5 illustrates how this study differs from the majority of previous social network research.

Table 2.4. Condensed summary of the literature on social networks.

Category	*Examples*	*Summary*
Theoretical	Social capital (Bourdieu 2011; Coleman 1988; Lin 1999), social comparison (Festinger, 1954), structural holes (Burt, 2002), strength of weak ties (Granovetter, 1983).	*Substantial application in offline networks. Limited application in online networks, unclear if these theories generalize to online networks** (Kane et al., 2012).
Descriptive	Network size, network connectivity, power law distribution of ties (Newman, 2001).	Mature body of research in both online and offline settings.
Impact	Adoption (Benbasat and Wang 2005), virtual team performance (Ahuja & Carley, 1998; Kanawattanachai & Yoo, 2007), job performance (Merten & Gloor, 2010; Zhang & Venkatesh, 2013), job satisfaction (Gloor 2016)	Substantial evidence that both online and offline networks are consequential in impact in both organizations and societies.
Contagion	Happiness (Fowler and Christakis 2008), smoking (Christakis & Fowler, 2008), product endorsement (Carmi et al., 2012).	Substantial evidence for offline network contagion. Inconclusive but ongoing work in online settings.
Formation	Structural factors e.g. preferential attachment (Barabási & Bonabeau, 2003), reciprocity (Bapna et al., 2011), transitivity (Agneessens & Wittek, 2008; Block, 2015), personal (demographic) characteristics such as racial homophily (McGuire, 2002), gender homophily (Cooper, 1997), genetics (Fowler et al., 2009) male-female sociality (Goodreau, Kitts, & Morris, 2009)	Evidence that structural factors matter in formation of offline (Van Duijn, Zeggelink, Huisman, Stokman, & Wasseur, 2003) and online networks (Capocci et al., 2006; Johnson et al., 2014), *but little work exists because of difficulty in modeling structural dependencies among ties, and correlations among structural factors* (Harris, 2013; Morris, Handcock, & Hunter, 2008). New statistical tools may address these difficulties. *Very little work done in organization settings.*

*Italicized text shows the understudied areas of social network research addressed by this study.

Social contagion provides a telling example on the importance of studying network formation. In a network in which tie formation is mainly a result of homophily, the network may be segregated across gender, racial, and education lines; hence, for a variety of phenomena, identifying the criteria by which ties form may help target efforts to encourage or discourage the spread of those phenomena. If income is a predictor of tie

formation in a particular network, then negative specific phenomena may spread faster within particular income brackets compared to others.

Table 2.5. Differences between this study and previous research

Existing Research	*This study*
Effects of social networks	Antecedents of social networks
Static snapshots of networks	Longitudinal
Single network at a time	Multiple networks
Descriptive	Statistical

Thus, social network formation can avail more granulated explanations of phenomena, and in the process, may avail evidence for more targeted interventions in the case of negative phenomena. Some examples in the literature are illustrative. Happiness, obesity, smoking etc. have been shown to spread throughout offline social networks by the contagion mechanism (Christakis and Fowler 2007, 2008, 2013; Fowler and Christakis 2008; Kramer, Guillory, and Hancock 2014). In other words, individuals that are connected through some relationship to someone that is happy are more likely to be happy than individuals that are not; the same is true for obesity and smoking. However, what connects these phenomena is the nature of the networks facilitating the spread of the phenomena is that they are offline e.g. friendship in the spread of smoking behavior and obesity (Christakis and Fowler 2007, 2008). However, evidence is now accumulating that contagion also exists in online networks, and the implications are profound.

Studies of contagion in online networks show that face-to-face interactions are not a necessary precondition for influence to spread from person to person. Depression and happiness can spread from person to person by simple exposure to information on online

platforms (Kramer, Guillory, and Hancock 2014; M. A. Moreno et al. 2011). For illustrative purposes, let us explore the widely reported Facebook experiment of 2014, where researchers exposed different groups of users to varying amounts of positive and negative posts in their news feeds (Kramer, Guillory, and Hancock 2014). Users exposed to fewer positive posts subsequently produced fewer positive posts and more negative posts; and the reverse was also true i.e. those exposed to more positive posts produced more positive posts and fewer negative posts. The findings of the study are evidence that positive and negative emotions can be spread to others even *without* in-person interaction. These findings have some important potential implications, chief among which is the ability to influence people's emotions at scales that were previously unimaginable.

To reiterate, network formation is not random, but rather it materializes out of certain structural tendencies and demographic influences. In political elections, demographic influences are particularly salient; in the United States (US) for example, black voters tend to vote differently than white voters, women tend to vote differently than men, young people vote differently than older people etc. (Jackman and Vavreck 2010). If online social networks exhibit racial, gender, and age homophily respectively, political strategists can expose different messages to a limited number of people within specific groups on these online networks, potentially saving lots in spending in political advertising. The savings arise from information diffusing through the networks via online social network contagion; rather than broadcasting information to everyone within the network, advertisers need only expose salient messages to a subset of their target demographics, contagion takes care of the rest. Similar reasoning can be applied for other

advertising domains where demographic targeting is relevant e.g. television shows, clothes, automobiles, credit cards etc.

In organizations, positive and negative messages diffuse throughout the organization over face-to-face, email, and other social media channels. Such diffusion can have implications for employee morale; rumors about impending bonuses and salary raises boost morale, and rumors about layoffs depress morale (Bordia et al. 2006; DiFonzo, Bordia, and Rosnow 1994). Using email, such rumors may propagate fast throughout organizations. In fact, rumors materialize in different forms to different subnetworks (DiFonzo et al. 2013). In organization networks characterized by homophily and/or other demographic differences in network position, news or rumors about emerging project and promotion opportunities may spread faster in particular subnetworks, and slower in others. Homophily means that people sharing certain traits form more closely connected subnetworks than others, and such subnetworks will likely receive different versions of rumors e.g. the false rumor that *Tropical Fantasy Fruit Punch* was owned by the Ku Klux Klan initially propagated just to African American subnetworks, and a rumor about a boy found mutilated in a Detroit mall identified the victim as white in white subnetworks, and as black in black subnetworks (DiFonzo et al. 2013; Fine and Turner 2001). Because rumors may deepen or alleviate uncertainties in different organizational subnetworks, tie formation patterns implicate variations in employee morale according to demography (DiFonzo et al. 2013; DiFonzo and Bordia 2007). These implications for the spread of information and sentiment underscore the importance of studying online network formation.

Summary of Chapter Two: Background Research

In this chapter, we introduced social networking concepts and outlined several characteristics of the features that distinguish one network from another. Next, we recounted a brief history of social network research. We next described how the social network perspective is used in organizational studies and IS research. Throughout these descriptions, we highlighted how online networks may help in answering hitherto difficult questions in social network research e.g. easier and perhaps more accurate measurement of centrality scores, which in turn might be more informative than is presently understood and understanding the antecedents of centrality. We concluded by pointing out the opportunities offered by online networks in answering certain questions e.g. questions on network formation and on potential differential effects of social contagion on different demographic groups within organizations.

In the next chapter, we outline some hypotheses that should help reveal answers to our research questions.

CHAPTER THREE

Theoretical Foundation and Hypothesis Development

Factors Affecting the Formation of Online Social Networks

Earlier work on social networks has implicated three mechanisms in social network formation – preferential attachment, homophily, and, in the case of directed networks, reciprocity (Kilduff and Tsai 2003; Schaefer et al. 2010; Wasserman and Faust 1994). Preferential attachment is a tendency for nodes to join a network by attaching to popular nodes (Barabási and Bonabeau 2003). Homophily is the tendency for nodes to prefer connecting to nodes with similar characteristics (McPherson et al. 2001). And reciprocity is the tendency for nodes with an incoming tie from one node to have an outgoing tie to that node (Wasserman and Faust 1994). However, we identify another feature, a tendency towards high density of ties as a key feature of online social networks. This is because online networks facilitate convenient and affordable interactions among organization members, even those separated by geographic distance and power (Kane et al. 2012; Klein et al. 2004). Therefore, it is likely that online networks will display higher density than offline networks.

In this section, we examine the identified tendencies of centralization, homophily, reciprocity, and isolation under two competing social network theories. Existing social network theories have been generated from studies on offline networks (Kane et al. 2012). However, electronic tools of communication are changing the nature of social interactions, which suggests that existing theories might prove inadequate in explaining

online social networks (Agrawal et al. 2011). Hence, it is possible that online social networks might require multiple theories or even new theory to explain them. In that spirit, this study employs two theories to create two competing models of network evolution. The theories are social capital theory and social comparison theory and they are described in the following subsections.

Social Capital Theory

Social capital theory postulates that an individual acquires access to resources through building social relationships with others (Agneessens and Wittek 2008; Bourdieu 2011; Coleman 1988; Kilduff and Tsai 2003). Embedded in social networks are important features of social organization such as trust, norms, and sanctions for violating those norms (Burt 2002; Coleman 1988). Social relationships grant access to organizational resources through three mechanisms: information, influence, and social credentials (Lin 1999). First, an individual acquires valuable information from his/her social connections. Such information includes news about promotion opportunities, projects, and impending retrenchments. Second, a relationship with influential members of the organization allows one to exert some indirect influence on organizational decisions. Third, the acquisition of social ties helps one acquire social credentials which are certifications of one's social tie resources (Bourdieu 2011; Lin 1999). As such, relationships can be very valuable in an organization. Building and maintaining a relationship is costly, hence individuals have a limit on the number of relationships they can sustain (Kilduff and Tsai 2003). Further, the movement of resources is generally expected to be bidirectional if the relationship is to be sustained. These cost

considerations and the desire to use relationships as a key to organizational resources make the social capital perspective an instrumental rather than emotional one.

The following three points summarize the social capital perspective and will be instrumental in building the hypotheses:

(i) Actors acquire access to resources by building and maintaining relationships with others.
(ii) Not all relationships are the same. Some network nodes are more influential than others; hence, relationships with such nodes are more valuable in social capital terms.
(iii) An individual's social capital can be depleted. Flouting organizational norms is one way in which an individual's social capital can be reduced.

Social Comparison Theory

Social comparison theory aims to explain how individuals make choices as to whom to interact with. The theory was originally formulated to explain how individuals evaluate their opinions and abilities relative to others (Festinger 1954; Kilduff and Tsai 2003). Humans possess a strong innate drive to evaluate their opinions and abilities (Festinger 1954). To accomplish this evaluation, the individuals compare their opinions or abilities to those of similar others (Gibbons 1992). For example, when individuals assess the values of their opinions concerning some political development, they are likely to compare those opinions to those of fellow political ideologues. This preference for similar others ultimately extends to the choice of whom to interact with. Hence, when faced with the choice to interact with two other people, individuals will interact with the person with whom they are more similar. This tendency to prefer similar others is called homophily, and it stems directly from social comparison theory (McPherson, Smith-Lovin, and Cook 2001). For any individual, multiple criteria can be used as a basis for categorization of others e.g. age, gender, race. However, past research has shown that

when conflict arises among criteria, individuals will choose the relatively rare criterion to categorize themselves as the basis for relationship formation (Mehra et al. 1998). For example, in an organization dominated by one racial group, an individual in a racial minority may use race, rather than gender, as the criterion of categorization when selecting communication partners. Similarity criteria also extends beyond demographic characteristics to network-specific criteria (Festinger 1954; McPherson, Smith-Lovin, and Cook 2001; Contractor and Eisenberg 1990). Hence, given a choice between two individuals, a person will choose to interact with the individual with a similar personal network.

The following points summarize the social comparison perspective and they will be used when building the hypotheses:

(i) Individuals select others to communicate with based on their similarities (homophily). Similarities can be demographic characteristics such as age, gender, and personality or network-specific characteristics such as popularity and influence.

(ii) Individuals possess the discretion to choose the most important criterion in their decisions on whom to interact with. The rarer the criterion, the more likely it is to be used as a basis for relationship formation.

The social comparison perspective predicts that, in an email communications network, network participants are likely to select others to maintain email relationships with based on certain characteristics. The more similar a potential connection is, the greater the probability than an individual will form a communication relationship with that connection.

Table 3.1 summarizes the differences between social capital theory and social comparison theory.

Table 3.1. Differences between social capital and social comparison theories.

Dimension	*Social capital*	*Social comparison*
Reason for tie formation	People create ties to gain access to resources	People create ties to interact with similar others.
Nature of motivation to form tie	Instrumental/economic	Affective/emotional
Predicted network structure	Centralized allocation of influence	Egalitarian allocation of influence

Next, we use the two theories to build competing hypotheses concerning tie formation in online networks.

Centralization in Online Networks

The social comparison and social capital perspectives predict different outcomes on centralization within an online social network. Under the social capital perspective, individuals form relationships to gain access to resources. Because most organizations are organized in hierarchical fashion, access to resources is also non-uniformly distributed towards organization members (Fischer 2004). Hence, chief executives have higher access to resources than middle managers, and supervisors than their subordinates. The social capital perspective predicts that individuals would prefer to form online ties with people that have greater access to those resources than with people with less access. Further, online networks mediate the distribution of valuable information within the organization (Agrawal et al. 2011). Examples of valuable information include information about promotion opportunities, new project opportunities, retrenchment, and general gossip (Lin 1999; Kane et al. 2012). Such valuable information is initially available to few within the organization; hence, forming email ties with such connected individuals can be advantageous. The hierarchical structure that characterizes most

organizations resembles a pyramid i.e. very few people at the top and the majority at the bottom. As such, the distribution of access to information is also likely to follow this shape, with the highest ranked people enjoying the greatest access to valuable information and the least ranked having the least access. Over time, information on the identities of the most influential individuals within the network percolates throughout the organization, ensuring that centralization strengthens even more. As such, the social capital (SC) perspective postulates that:

H1 (SC): Over time, an online network evolves with a tendency towards centralization.

Social comparison, on the other hand, predicts a democratization of access to information over time and decentralization of the network over time. This is because people create relationships on the basis of similarity. Hence, men prefer forming online relationships with other men, women with other women, and lowly ranking individuals with other lowly ranking individuals etc. Ultimately, individuals will find people to connect with that they share similar attributes with, implying that access to resources weakens as a factor determining the formation of relationships. In other words, individuals form relationships for reasons other than the instrumental access to resources. In an organization, this means relationships are formed on the basis of forming and/or maintaining friendships, seeking or sharing advice, invitations to social events, sharing jokes and memes etc. As such, online relationships are likely to be uniformly distributed among organizational members. Over time, individuals acquire more opportunities to interact with similar organization members, ensuring that relationships accrue even more uniformly to everyone, thus effectively flattening the hierarchical structure that

characterizes the formal organization. The social comparison (SCMP) perspective hence predicts that:

H1 (SCMP): Over time, an online network evolves with a tendency towards decentralization.

Reciprocity in Online Networks

The concept of reciprocity is captured in the well-known idiom "one good turn deserves another" (Cohen and Bradford 2007). The normative view of reciprocity is particularly relevant in the social capital perspective. Individuals build social capital not only by creating relationships, but by maintaining them as well. This means that invitations to connect with others are likely to be accepted and reciprocated under the social capital perspective, otherwise the opportunity to acquire social capital is lost. Further, reciprocal relationships are beneficial in the long run because they facilitate more efficient knowledge acquisition (Chiu, Hsu, and Wang 2006). Individuals that ignore invitations to form communication relationships also run the risk of depleting their social capitals within the organization, hence the social capital perspective suggests that individuals will reciprocate overtures to form relationships with others.

H2 (SC): Over time, an online network evolves with a tendency towards reciprocity.

Social comparison predicts a similar outcome. In labor relations, generous wage offers by employers are reciprocated with higher effort by employees; in this regard, employees give back to the employer levels of effort commensurate with their wages (Gächter, Nosenzo, and Sefton 2012). Similarly, in online communication, individuals compare their communication behaviors with that of others. If they perceive that others

are exerting great effort to form ties with them, they are also likely to exert similar levels of effort, to narrow the difference in perceived effort. Over time, this process repeats as relationships grow stronger, leading to even greater levels of reciprocation. Hence:

H2 (SCMP): Over time, an online network evolves with a tendency towards reciprocity.

Density in Online Networks

The social capital perspective predicts an increase in network density over time. An employee gains access to resources by building communication relationships with others. Such resources include, at the very least, materials and information necessary to accomplish work-related tasks. Every organizational member requires access to resources, and likely maintains a relationship with at least one other organization member in order to perform their tasks. In other words, according to the social network perspective, every organization member is useful and eventually finds other people to connect to. The network becomes denser as organizational members become more aware of each other over time and form communication ties to establish formal and informal cooperative endeavors.

H3 (SC): Over time, an online network evolves with a tendency towards high density of ties.

In the social comparison perspective, individuals form connections with similar others. In an organization, it is likely that some individuals may have difficulty finding similar others, according to the criteria they attach importance to. In other words, each categorization creates minorities, and as the number of such categorizations increase: gender, education, age etc. the probability that some nodes are isolated will rise. When

individuals connect on such bases, they tend to not seek ties with people outside their own groups (Ibarra 1992; McPherson, Smith-Lovin, and Cook 2001). Individuals in minority categories thus struggle to form ties. The network thus fractures along demographic lines, reducing the overall density of the network; hence, in the social comparison perspective:

> H3 (SCMP): Over time, an online network evolves with a tendency towards low density of ties.

The Effect of Education on Tie Formation in Online Networks

Education is valuable within an organization, because it is a source of knowledge and expertise (Klein et al. 2004). Hence, social capital predicts that educated individuals will be popular within the network; other individuals will communicate with educated individuals to gain access to their specialized knowledge and expertise. Further, highly educated individuals may send fewer outgoing ties to lowly educated individuals (owing to the expertise discrepancy), hence widening the popularity difference between highly educated and lowly educated individuals. In organizations, email use is positively associated with education (Phillips and Reddie 2007); this implies that higher educated individuals have more email contacts than lower educated individuals. Education is also positively associated with status (Klein et al. 2004); hence highly educated individuals may be perceived to be of high status within the organization. This high status attracts more incoming ties, hence:

> H4 (SC): Over time, higher educated nodes are more popular than lower educated nodes in online networks.

The Effect of Gender on Tie Formation in Online Networks

Under social comparison, individuals with similar genders have greater affinity for each other. Although electronic tools offer fewer visual cues of potential contacts than face-to-face interactions, offline interactions may carry over to electronic platforms. This means that the homophily that is commonly observed in offline networks will likely be evident in online networks as well. Take women, for example, several studies show that they are more likely to form ties with each other than predicted by chance because of their shared struggles navigating work-life balance, combating sexual harassment, and their minority status in certain industries (McPherson, Smith-Lovin, and Cook 2001; Watts 2003). These shared experiences make them more likely to form online communication ties with each other:

> H5 (SCMP): Over time, ties are more likely to form between people of the same gender than between people of different genders in online networks.

The Effect of Personality on Tie Formation in Online Networks

Social comparison predicts that individuals with similar personalities are likely to form ties among each other. Introverts may prefer forming ties with other introverts, and online communication may enable them to find each other and communicate. Individuals prefer interacting with individuals with whom they are similar, and they look to such for evaluating their abilities and opinions (Festinger 1954). Relationships in offline networks demonstrate that introverts prefer forming ties with other introverts, and extroverts with other extroverts (Gattis et al. 2004; Feiler and Kleinbaum 2015). Individuals with similar personalities are more likely to be attracted to each other than individuals with different personalities, or they may encounter each other more often than individuals with different

personalities (Feiler and Kleinbaum 2015)In organizations, these offline ties may be sustained by using electronic communication tools such as email. Hence, introverts are likely to prefer forming online relationships with other introverts, and extroverts with extroverts. Thus, social comparison posits that online networks exhibit homophily according to personality:

H6 (SCMP) – Over time, ties are more likely to form between people of the same personality than between people of different personalities in online networks.

Table 3.2 shows the two sets of competing hypotheses from social capital theory and social comparison theory regarding an email communication network's preferential attachment, reciprocity, tendency towards isolates, and the effects of education, gender, and personality on tie formation.

Table 3.2. Competing hypotheses according to social capital and social comparison theories.

Hypothesis	*Social Capital Perspective*	*Social Comparison Perspective*
H1	Over time, an online network evolves with a tendency towards centralization.	Over time, an online network evolves with a tendency towards decentralization.
H2	Over time, an online network evolves with a tendency towards reciprocity.	Over time, an online network evolves with a tendency towards reciprocity.
H3	Over time, an online network evolves with a tendency towards high density of ties.	Over time, an online network evolves with a tendency towards low density of ties.
H4	Over time, higher educated nodes are more popular than lower educated nodes in online networks.	
H5		Over time, ties are more likely to form between people of the same gender than between people of different genders in online networks.
H6		Over time, ties are more likely to form between people of the same personality than between people of different personalities in online networks.

Differences Between Online and Offline Networks

Much of existing social network theory research utilized studies of offline networks such as friendship, advice, and marriage networks. Recently, however, online networks have become more ubiquitous and more powerful (Agarwal, Gupta, and Kraut 2008; Muchnik et al. 2013; Oinas-Kukkonen, Lyytinen, and Yoo 2010). To the extent that they facilitate relationships among people, online networks are similar to offline networks. Yet, at least three fundamental differences exist. First, electronic tools afford individuals expanded possibilities to interact with each other (Kane et al. 2012); this likely creates denser networks than are possible with face-to-face interactions. Second, online networks are characterized mainly by transitory and more effortless exchanges of information; for example, to become friends with another person on Facebook, one needs only to click a few buttons and type in a few letters of text at most, and on Twitter one may similarly follow another individual. It is not clear, therefore, what a concrete definition of an online relationship should be. Third, online networks aggregate the functions of more than one network at a time. On Facebook, as an example, people that are "friends" might have different relationships that actually define them, like family, work, mutual organization membership relationships etc. Hence, what is defined as a friendship relationship might actually be a different kind of relationship. Nevertheless, traditional social network theories offer a good starting point for understanding the nature of online social networks, and below we propose that levels of centralization, reciprocity, density, and homophily likely differ in online and offline networks. We also argue that an individual's position within offline networks likely partially determine their positions within online networks.

Centralization

The effect of using electronic tools on organization networks is likely complex. On one hand, electronic tools potentially flatten social hierarchies by destroying social boundaries; less powerful individuals can communicate and interact with more powerful individuals (Klein et al. 2004). These flattened social hierarchies potentially weaken centralization of online networks. Further, a network's level of centralization is likely to be affected by the visibility of its most popular individuals (Newman 2001b). Influential individuals are likely more visible in offline networks than in online networks. For example, the person to whom most employees turn to for work related advice may be visible to everyone within a single department; in contrast, the employee that communicates the most with everyone else using email might be less visible. Other physical cues that make one popular are also more prominent in offline networks (Fowler, Dawes, and Christakis 2009); for example, two individuals with similar expertise may have different levels of politeness. The more polite individual might then draw even more connections than the less polite individual. Hence, offline interactions generate more cues that enhance centralization than online interactions. This visibility disparity likely causes differences in levels of centralization between offline and online networks. Because the most influential individuals in offline networks are likely well-known in the organization, other employees are likely to initiate ties to them because of their popularity. For example, individuals that are renowned for their technical skills may increasingly acquire new ties in the advice network.

On the other hand, the use of electronic tools may entrench existing hierarchical structures within the organization. Individuals known for their access to novel and

relevant information are likely to draw disproportionately more incoming ties than the rest of the network. Individuals that are prominent for their technical expertise, for example, may become even more popular within an email network because people removed from them by geographical distance and other barriers may more easily communicate with them. The easier effort it takes to establish communication ties with prominent individuals online makes it more likely that they should receive even more incoming ties. Online social media such as Facebook and Twitter demonstrate this effect, already prominent people such as politicians and musicians are easily the most popular people on these platforms, and hence they replicate their offline success onto online networks[1]. Thus, we hypothesize:

H7: Centralization is greater in online networks than in offline networks.

Reciprocity

The consequences of unreciprocated ties are likely to be greater in offline networks than in online networks for a number of reasons. First, tie formation is likely more visible in offline networks than in online networks. Take an advice network, if one only asks for advice without giving back that person may garner a reputation for selfishness. Further, the effort required to initiate an offline relationship is high compared to the effort required for online relationships, meaning that individuals would find it easier to reciprocate such relationships. Individuals are also much more likely to remember their offline interactions than their online interactions because the former are likely to be physical encounters. Hence, to avoid depleting their social capitals, individuals are more likely to reciprocate in offline networks. Online networks, on the

[1] http://friendorfollow.com/twitter/most-followers/

other hand, hide their tie connection requests, such that individuals likely suffer limited consequences from rejecting tie requests. Further, because it takes relatively easy effort to send tie requests in online networks compared to offline networks, ignoring these requests in online networks may incur lower penalties. Hence:

H8: Reciprocity is greater in offline networks than in online networks.

Density

Offline relationships require more effort creating and maintaining them than online networks (Kane et al. 2012). With face-to-face communication, individuals have to be physically co-present, which limits the number of relationships that any one individual can participate in. Therefore, it is likely that certain members of the organization will have no ties within offline networks. On the other hand, electronic tools require relatively little effort to compose and send messages to other organization members; for example, Gmail generates an option to respond to an RSVP message with just a single click. Further, electronic tools make it easier to find potential connections across spatial and social boundaries; e.g. individuals in different offices of the same company can communicate easier with email, and junior level employees can contact senior management using email. In this regard, online communication makes it more likely for every individual to be connected to at least one other person within the network. The major consequence of this effort disparity is a higher density for offline networks relative to online networks.

H9: Online networks are denser than offline networks.

Homophily

Individuals are more visible to each other in offline networks than in online networks. When people interact face-to-face, they are able to assign gender, ethnicity, personality values etc. to each other. Certain electronic tools such as email and social media may also reveal certain attributes such as gender, age, and ethnicity if individuals provide the permissions to do so. At the same time, individuals may decide to submit false information e.g. false date of birth and profile pictures belonging to other people. Further, it is highly probable that individuals will inaccurately draw conclusions as to what demography an unfamiliar person belongs to, because profile pictures are static and may convey inaccurate cues. Hence, if individuals prefer interacting with people of similar characteristics, the cues that enable the discovery of such characteristics are more prominent in offline networks than in online networks. Overall, therefore, observed homophily should be greater in offline networks than online networks:

H10: Homophily is greater in offline networks than in online networks.

The Effect of Offline Network Position on Tie Formation in Online Networks

Other networks within an organization are likely to influence its online networks. This is because electronic tools sometimes subsume the core functionalities of other networks. For example, individuals sometimes use email to seek advice from their co-workers, and to send instructions to their subordinates; hence, individuals' positions in the email network are likely to be influenced by their positions within advice and working networks. Even then, the impact of offline networks on online networks is likely to be complex. Online networks may complement, echo, or even substitute existing offline networks (Kane et al. 2012). Influential individuals in the advice network are

likely to be popular within the online network. As individuals acquire reputations for solving specific work-related problems, they are also likely to receive more emails with solicitations for such expertise. Individuals in the brainstorming network, on the other hand, are likely to use more interactive methods of communication in their brainstorming sessions; hence, face-to-face interactions might be more useful in forming the brainstorming network. Ultimately, as the primary nature of communication required to form a particular network varies from asynchronous to synchronous, that network is less likely to echo the online networks within the organization. As such:

> H11 – The greater the synchronicity required in the activity underlying an offline network, the weaker the relationship between an individual's position within that network and his/her position in an online network.

Table 3.3 shows the hypotheses that capture the differences among online and offline social networks, as well as the influence of offline network centrality on an individual's position within the online network.

Table 3.3. Differences and interplay between online and offline networks.

Hypothesis	*Description*
H7	Centralization is greater in online networks than in offline networks.
H8	Reciprocity is greater in offline networks than in online networks.
H9	Online networks are denser than offline networks.
H10	Homophily is greater in offline networks than in online networks.
H11	The greater the synchronicity required in the activity underlying an offline network, the weaker the relationship between an individual's position within that network and his/her position in an online network.

CHAPTER FOUR

Research Methodology

In Chapter Three, we developed and outlined some hypotheses regarding the formation of online social networks and the differences between offline and online social networks. This chapter describes the research context and outlines the methods of data collection and data analyses required to examine these hypotheses.

Research Context

The site of data collection is a company (pseudonym EmailCo) that specializes in data analytics. The company employs between 40 and 50 people at any single point in time, who are based in Singapore (headquarters), the US, the UK, and the Philippines among other countries. We included a total of 41 people that were consistently part of the organization during the period of analysis, consistent with prior research on social networks (Agneessens and Wittek 2008). Although 41 is a small sample size for traditional survey-based studies, it falls within the typical range for organizational social network research; examples are n=32 in Yuan and Gay (2006)'s study, n = 42 in Thaden and Rotolo (2009)'s study, and n=87 in Sykes et al. (2014)'s study. EmailCo is organized into eight different departments: technology, administration, leadership, sales, labs, operations, professional services, and strategy. It stores its email communication logs which we have access to. We have access to data spanning the last three years. The company also granted us access to an attribute file containing the gender, department, and organization rank of each employee.

Data Collection

Our data collection proceeded in two phases – online network and offline network. We recount these processes below.

Phase One: Collection of Online Network Data

EmailCo maintains a record of metadata for every email sent or received by an employee. For this study, we use records dating from July 2015 to August 2017. The metadata included "from" and "to" fields along with a timestamp for each email transaction. Each employee in the organization has a real email address and an associated pseudonym email address. The "from" and "to" fields contained the pseudonymized email addresses for each email transaction. We filtered the data to exclude any external communications and retain only internal email transactions i.e. those involving only EmailCo employees. Some of the internal email addresses were bots e.g. "admin@emailco.com" and "postmaster@emailco.com"; we similarly removed all email transactions involving these bots from the data sample.

We filtered the data by month and obtained 26 separate files for each month; afterward, we constructed 26 networks corresponding to each of the months. To construct the networks, we carried out the following steps for each month:

1. Create a comma separated values (csv) file with 3 columns: "From", "To", and "Count".
2. For each pair of email addresses i and j in the "From" and "To" columns, aggregate the number of emails sent from i to j and record it in the "Count" column. Thus, "Count" is the number of emails sent from i to j in one month.
3. The networks are directed; hence we repeat (2) for every j to i.

4. Calculate the median value for the "Count" column. For an outgoing tie to exist from *i* to *j*, the "Count" value must match or exceed the median value. Dichotomizing ties is common in the social networks literature (e.g. Hanneman and Riddle 2005; Sykes, Venkatesh, and Gosain 2009).
5. Delete the "To" and "Count" values for all rows with "Count" less than the median calculated in (4). We retain the "From" value to identify the isolates in the network.
6. Delete the rows for all self-loops, or rows for which the "From" and "To" values are identical.
7. The output from (5) is called an edgelist (ties are also called edges in SNA), and it contains the ties that qualify for inclusion in the network.
8. Export the edgelist into *R* (Ihaka and Gentleman 1996) and carry out all network related analysis using the *R statnet* package (Handcock et al. 2008).

In Phase 1 we collected online network data. Phase 2 was for collection of offline network information.

Phase Two: Collection of Offline Network, Personality, and Demographic Data

Networks. We obtained offline network information via a survey using generally accepted sociometric techniques (Sykes, Venkatesh, and Johnson 2014; Wasserman and Faust 1994). We administered the survey electronically via the Qualtrics platform. EmailCo provided us with a mapping of real email addresses and pseudonymized email addresses. We sent the company's 46 employees a link to the survey via email. Data collection took about a month. We sent two reminder emails to employees that had not responded after a week and three weeks respectively. The respondent rate was 41/46, or

89%, which exceeds the 80% threshold specified in prior social network studies (Zhang and Venkatesh 2013).

We selected three networks as representative of traditional offline networks[1]. These networks were advice, brainstorming, and working networks at play within the organization. The advice network represents the configuration of expertise-seeking ties of employees within the organization, and it captures who seeks advice from whom in carrying out work-related activities. To capture advice seeking ties in the network, we asked respondents to "*indicate the extent to which (they) turn to each of the following people for expert advice about work-related activities*."

The brainstorming network represents the configuration of idea generation and problem-solving ties of employees within the organization, and it captures who brainstorms or innovates together with whom. The prompt for determining whether an advice tie exists is the statement "*indicate the extent to which you innovate or brainstorm together with the following people*."

The working network is the configuration of everyday work ties of employees within the organization, and it captures who works with whom on a day-to-day basis. To establish whether an advice tie exists between any two people, we asked participants to "*indicate the extent to which (they) work together with the following people*."

To construct the networks, we displayed, for each respondent, a full roster of the 46 employees within the organization and prompted them with statements to assess the

[1] There is some nuance here. Advice networks have traditionally been identified using offline methods of data collection i.e. surveys, field observations, interviews etc. It is possible that people may seek advice using electronic means such as email or electronic messaging, but this is a question we seek to answer, do electronic networks echo what has traditionally been observed in offline networks?

strengths of their relationships with the other employees. Consistent with prior methods of collecting social network construction data, the responses were on a 5-point scale (never, rarely, somewhat often, often, very often) (Sykes, Venkatesh, and Johnson 2014). To minimize the effort burden on respondents, the survey stipulated that respondents only indicate a response for another employee if they perceived a relationship to exist with that employee i.e. if the extent matched or exceeded "rarely". An outgoing tie was deemed present if the extent of interaction matched or exceeded "somewhat often"; dichotomizing the tie strengths in this manner is also consistent with prior social network research, and is necessary for examining tie formation using exponential random graph modeling (ERGM) (Cross, Borgatti, and Parker 2001; Harris 2013; Sykes, Venkatesh, and Gosain 2009).

Using the responses gathered from the survey, we created three 41x41 matrices corresponding to the advice, brainstorming, and working networks, with a "0" or "1" between every pair of employees *i* and *j*; "1" indicated the existence of a tie and "0" its absence. As with the email network, directionality mattered; for example, *i* might indicate that s/he often seeks advice from *j*, but *j* might indicate that s/he never seeks advice from *i*; in that case the intersection *i,j* will have a value of 1, but the intersection *j,i* will have a value of 0. We removed all self-loops when analyzing the network i.e. replacing all *i*,*i* non-zero values with a "0". Finally, we exported the three matrices[2] into *R* for further analyses using the *statnet* package.

Hypothesis H11 examined the effect of offline network centrality on online network formation. Hence, we obtained popularity, gregariousness, betweenness, and

[2] Matrices and edgelists are equivalent structures and they convey the same information. A survey generates matrices.

eigenvector centrality scores for each individual in the three offline networks using *R*'s *igraph* package (Csardi and Nepusz 2006). We added these centrality scores to the attribute file.

Personality. The survey also contained questions to assess the extraversion/introversion personality characteristics of respondents. There were 12 questions, obtained from and validated in prior research (Sato 2005). This measure of extraversion/introversion has the advantage of brevity relative to other measures in the literature (e.g. Eysenck and Eysenck 1965), and therefore lightens the burden on respondents. We also asked employees to indicate their genders and levels of highest education.

Control variables. We asked respondents a series of questions to control for the influence of other variables on network formation. The first variable is the number of known EmailCo contacts prior to joining. Individuals with a high number of prior contacts might be more popular than individuals with few contacts because they likely start their EmailCo experience with a high number of ties, and previous research shows that prior contact increases the probability of tie formation (Hahn, Moon, and Zhang 2008).

The second control variable is the extent to which employees assign importance to the use of email in conducting their work. Individuals that use email more often are likely to have greater numbers of email network ties than individuals that use email sparingly.

The extent to which employees assign importance to face-to-face interactions in conducting their work is the third control variable. We controlled for this variable

because individuals that rely mainly on face-to-face interactions may have more ties in offline networks than in the online network.

The fourth control variable is the extent to which employees assign importance to phone conversations when conducting their work activities. It was important to control for this variable because individuals that mainly use the phone might use email sparingly, hence decreasing their email network tie numbers.

We also controlled for task interdependence and location of participants. Task interdependence is the extent to which team members require interaction and coordination in order to complete tasks (Guzzo and Shea 1992). Individuals with high levels of task interdependence might interact more with their team members using email than individuals with low task interdependence. We controlled for location because individuals might email their co-located workmates less than those that work at different locations. Table 4.1 shows all the variables that we used in the study.

Table 4.1. Study variables.

Variable	*Type of Variable*	*Date Range*	*Data Source*	*Fraction Missing*
Online network	Dependent	Jul-15 – Aug-17	Email Logs	0
Advice network	Dependent	Jul-17	Survey	0.05
Brainstorming network	Dependent	Jul-17	Survey	0.05
Working network	Dependent	Jul-17	Survey	0.05
Gender	Independent	Jul-17	Survey	0
Education	Independent	Jul-17	Survey	0.05
Personality	Independent	Jul-17	Survey	0
Number of prior contacts	Control	Jul-17	Survey	0.05
Task interdependence	Control	Jul-17	Survey	0.05
Email importance	Control	Jul-17	Survey	0.05
Face-to-face importance	Control	Jul-17	Survey	0.05
Phone importance	Control	Jul-17	Survey	0.05
Location	Control	Jul-17	Attribute File	0.17

Variable	*Type of Variable*	*Date Range*	*Data Source*	*Fraction Missing*
Organizational commitment	Dependent (post-hoc analysis)	Jul-17	Survey	0.05

Table 4.2 summarizes the characteristics of the sample.

Table 4.2. Sample characteristics

Attribute	*Category*	*Percentage*
Gender	Male	68%
	Female	32%
Race	White	37%
	South Asian	22%
	East Asian	15%
	Other	3%
	Undisclosed	22%
Education	Master's	22%
	Bachelor's	54%
	Some college	2%
	Associate's	12%
	High School	10%
Location	Headquarters	41%
	Non-headquarters	46%
	Unknown	13%
	Mean	*Std. Deviation*
Extraversion	3.39	0.77
Task interdependence	5.65	0.96
Organizational commitment	5.65	0.85
Email importance	4.34	0.72
Face-to-face importance	3.82	1.26
Phone importance	3.87	1.12
Number of prior contacts	1.87	2.85

Data Analysis

Exponential Random Graph Modeling

To examine our hypotheses, we employed exponential random graph modeling, a method that examines network formation. The method determines whether structural tendencies and demographic influences are statistically significant within a network (Morris, Handcock, and Hunter 2008). ERGM is needed because standard statistical models assume that observations are independent, but with ERGM one can model complex dependences among ties, because the ultimate goal in modeling social network formation is determining the factors that contribute towards network tie formation (ties are the basic building blocks of networks) (Robins and Morris 2007). The ultimate output of testing a model with this method is a probability of a tie between two individuals, given their personal characteristics and characteristics of the network. To the extent that it calculates the probability of a tie i.e. whether a tie is present or not, ERGM is analogous to binary logistic regression (Harris 2013). One difference is that standard statistical methods like regression assume that observations are unrelated, but the outcome variable for ERGM – the ties among nodes – are related, by the very definition of the word "network" (Harris 2013; Wasserman and Faust 1994). These probabilities are derived from the parameters calculated by ERGM. These parameters are what maximize the probability of the observed network, given the number of nodes comprising the network.

Parameter Estimation Procedure

The intuition behind ERGM stems behind the observation that networks are not formed randomly. A variety of features occur consistently in real world networks e.g.

power law distribution of ties (Barabási and Bonabeau 2003; Newman 2001a), homophily according to gender, race, education etc. (Ibarra 1995; McPherson, Smith-Lovin, and Cook 2001), and clustering i.e. tendency for one's friends to become friends as well (Harris 2013; Morris, Handcock, and Hunter 2008; Newman 2001b). Given that such features occur consistently, and simultaneously, how can one disentangle the features to see which ones are dominant? ERGM estimates maximum likelihood estimates of the parameters for each of these features such that the greater the magnitude of the parameter, the more dominant the feature.

Given a network of size *n*, a number of tie configurations are possible. An example is an undirected network with 3 nodes, A, B, and C. Some possible networks for this situation are 1) a network with zero ties 2) a network where all the nodes share a tie 3) a network in which A and B are tied but C has no connection etc. (Figure 4.1). How likely are each of these networks? The answer lies partly on structural tendencies and node characteristics. The simplest structural characteristic is the tendency for tie formation in the network or the Bernoulli model (Erdős and Renyi 1959), and this tendency varies across networks. Networks that vary from the Bernoulli model arise out of non-random features, and the non-randomness is what makes the networks interesting to study. If this tendency is high, we might expect the saturated network (top-left) to be more likely than the empty network (top-right). Or if gender homophily is a feature of this network, and we know that A and B are female and C is male, then the bottom-left network might be more likely than every other network. As the size of the network rises, so does the number of possible networks, and this number rises exponentially with respect to *n*. A model tested by ERGM includes these structural tendencies and node

characteristic influence as network statistics or predictors and hypothesize that they exist more than is expected of chance (Morris, Handcock, and Hunter 2008).

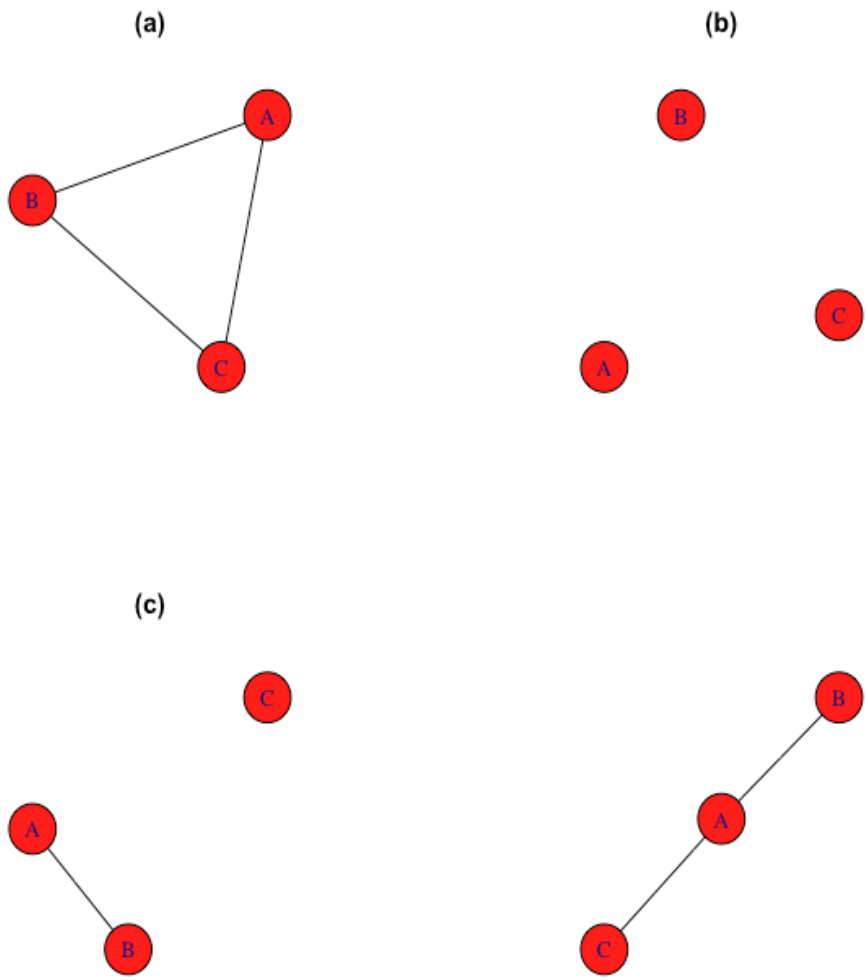

Figure 4.1. Possible network configurations for network with 3 nodes

ERGM generates the probability distribution of all the possible random networks of size *n*, the size of the observed networks (Morris, Handcock, and Hunter 2008). The probability of the observed network is calculated using the maximum-likelihood function:

$$P(Y=y)= \frac{\exp\{\theta^T z(y)\}}{\kappa(\theta, Y)}, \; y \in Y \quad (1)$$

Where:

- *z(y)* represents the set of effects e.g. reciprocity, popularity according to education, homophily according to gender, personality etc.
- θ^T - the vector of their associated weights representing statistical parameters.
- (θ, Υ) - the summed value of $\{\theta^T z(y)\}$ over all possible networks.
- κ - a normalizing constant to ensure that the total probability across all possible networks equals 1.

As the size of the network rises, the denominator $\kappa(\theta, \Upsilon)$ becomes infeasible to calculate, because the number of possible networks is exponential with respect to the number of nodes. Hence, ERGM uses Markov Chain Monte Carlo (MCMC) methods to estimate the probability distribution of all possible networks.

Equation (1) can be re-expressed in the following manner to obtain the conditional log-odds of a tie between two nodes:

$$\mathrm{logit}(\mathrm{P}(\mathrm{Y_{ij}} = 1 \,|\mathrm{n}, \mathrm{Y_{ij}^c})) = \sum_{\mathrm{k=1}}^{\mathrm{K}} \theta_{\mathrm{k}} \delta \mathrm{z_k}(\mathrm{y})$$

where:

- k is the number of network statistics in the model
- θ_k is the coefficient for each network statistic
- $\delta z_k(y)$ is the change in the network statistic when Y_{ij} from 0 to 1

The coefficient θ is the log-odds of a single tie between two nodes, conditional on the rest of the network (Harris 2013; Morris, Handcock, and Hunter 2008).

Specifications of Models

We employ two different models to predict tie formation in online networks. One model (Model 1) is the Bernoulli model (random ties) combined with the tendency towards centralization; this model tests H1. The other model (Model 2), which tests H2-H6, combines the Bernoulli model with reciprocity, homophily, and differential popularity (tendency for incoming ties) and gregariousness (tendency for outgoing ties)

according to gender, education, and personality, while controlling for popularity and gregariousness according to the number of prior contacts. We employed two different models because the more complex a model, the less likely it is to converge (Goodreau 2007), and in such cases one can employ multiple simpler models to identify the possible mechanisms through which the network comes to be (for examples see Goodreau 2007; Johnson, Faraj, and Kudaravalli 2014). We fitted all 26 networks corresponding to the months in the 26-month period of examination onto Model 1 and Model 2, to examine whether the factors influencing tie formation in online networks were consistent over the whole period.

To test hypotheses 7-10, we used Model 1 and Model 3. Model 3 is similar to Model 2 but is slightly altered to control for location and how respondents assign importance towards email, phone, and face-to-face interactions towards accomplishing their work. Model 3 also controlled for task interdependence. These controls were not included in Model 2 because their values might have changed over the course of the 26-month period over which we examined network evolution. The final model, Model 4, tested hypothesis H11, and included all the variables in Model 3 plus network centrality scores in the advice, working, and brainstorming networks.

Table 4.3 shows the descriptions of the models. Mathematical expressions of these models are found in Appendix B. We used the *R statnet* package test these four models. To estimate the parameters, the MCMC algorithm simulates a Markov chain of random networks constrained by the size of the observed networks; the algorithm generates a random network and compares its probability to that of the current network (Morris, Handcock, and Hunter 2008). The generated network is accepted as the next in

the chain if its probability exceeds that of the current network, otherwise a new network is generated. Table 4.4 shows the definitions of ERGM terms included in the models.

Table 4.3. Study models and descriptions

Model	*Description*	*Hypotheses tested*
Bernoulli	Random tie formation with tie probability equal to network density.	N/A (baseline model)
Model 1	Bernoulli model + tendency towards centralization	H1
Model 2	Bernoulli model + tendency towards reciprocity + demographic influences + *number of prior contacts**	H2 – H6
Model 3	Model 3 + *email importance* + *phone importance* + *face to face contact importance* + *location* + *task interdependence*	H7 – H10
Model 4	Model 4 + offline network centralities	H11

*Italicized variables are controls.

Table 4.4. ERGM term definitions adapted from Morris et al., (2008)

Term	*R Statnet Term*	*Description*
Edges	*edges*	The number of ties in the model relative to the potential number of ties. The Bernoulli model includes only the edges term.
Reciprocity	*mutual*	The number of ties for which a tie in the opposite direction exists i.e. for node *i* where *i* -> *j* and *j* -> *i* both exist.
Geometric weighted degree	*gwidegree*	This term is the sum of the values of the weighted degree distribution per node.
Homophily based on attribute *x*	*nodematch*	The number of ties for which two nodes share the same attribute value.

Term	*R Statnet Term*	*Description*
Main effect of a covariate *x*	*nodecov*	The sum of the values of continuous attribute *x* on node I and node j. *nodeicov* and *nodeocov* are the values on the sink (node with incoming tie) and source (node with outgoing tie) respectively. *nodeicov* measures popularity according to continuous attribute value, and *nodeocov* measures gregariousness according to continuous attribute value.
Main effect of a factor attribute *x*	*nodefactor*	For a factor attribute *x* with a number of unique levels, each level is assigned a value equaling the number of times a node with that level appears in a tie within the network. *nodeifactor* and *nodeofactor* only include values for which a sink and source with that level appear in a tie within the network respectively. *nodeifactor* measures popularity according to factor attribute level, and *nodeofactor* measures gregariousness according to factor attribute level.

In Chapter Five, we describe our findings.

CHAPTER FIVE

Results

Descriptive Statistics

Online Networks

We first describe the 26 networks corresponding to the months between July 2015 and August 2017 inclusive using 4 network characteristics – size, number of ties, density, and number of isolates (Table 5.1). The minimum network size was 45, the maximum was 59, the median was 48, and the mean was 48.9 with a standard deviation (SD) of 3.5. The number of ties in the network varied from 216 to 723, the median was 515 and the mean was 537.3, with a standard deviation of 98.5. The minimum network density was 0.11 (August 2016), the maximum was 0.32 (February 2016), the median was 0.23, and the mean was 0.23 with a SD of 0.04. Last, the number of isolates ranged from 0 to 11, the median was 4 and the mean was 4.231 with a SD of 2.76.

Figure 5.1 shows the changes in network characteristics over time. The number of nodes and number of isolates are both increasing in time, suggesting that as the company grows, it might be struggling to socialize its new members.

Online Network vs Offline Networks

The online (email) network was denser than the offline networks. Among the offline networks, the working network was the densest, followed by the working network, and lastly by the brainstorming network (Table 5.2).

Results from Exponential Random Graph Modeling

Parameters for network tendencies are summarized in Table 5.3. Full results for all the networks are in Appendix B, Tables B1 – B92. Positive parameter values indicate positive tendencies. The greater the absolute value of the parameter, the stronger it is.

Table 5.1. Network characteristics of online networks corresponding to the months between July 2015 and August 2017

Network	*Number of Nodes*	*Number of Edges*	*Density*	*Isolates*
Jul-15	47	508	0.24	1
Aug-15	45	476	0.24	0
Sep-15	46	534	0.26	1
Oct-15	47	498	0.23	5
Nov-15	45	465	0.23	5
Dec-15	47	489	0.23	3
Jan-16	45	483	0.24	3
Feb-16	48	723	0.32	4
Mar-16	48	485	0.22	5
Apr-16	47	484	0.22	4
May-16	48	628	0.28	0
Jun-16	49	514	0.22	4
Jul-16	52	711	0.27	0
Aug-16	59	666	0.19	7
Sep-16	56	638	0.21	8
Oct-16	54	623	0.22	11
Nov-16	48	592	0.26	5
Dec-16	48	484	0.22	7
Jan-17	45	216	0.11	7
Feb-17	49	543	0.23	5
Mar-17	49	571	0.24	4
Apr-17	49	541	0.23	4
May-17	51	561	0.22	5
Jun-17	53	515	0.19	8
Jul-17	50	513	0.21	3
Aug-17	46	508	0.25	1

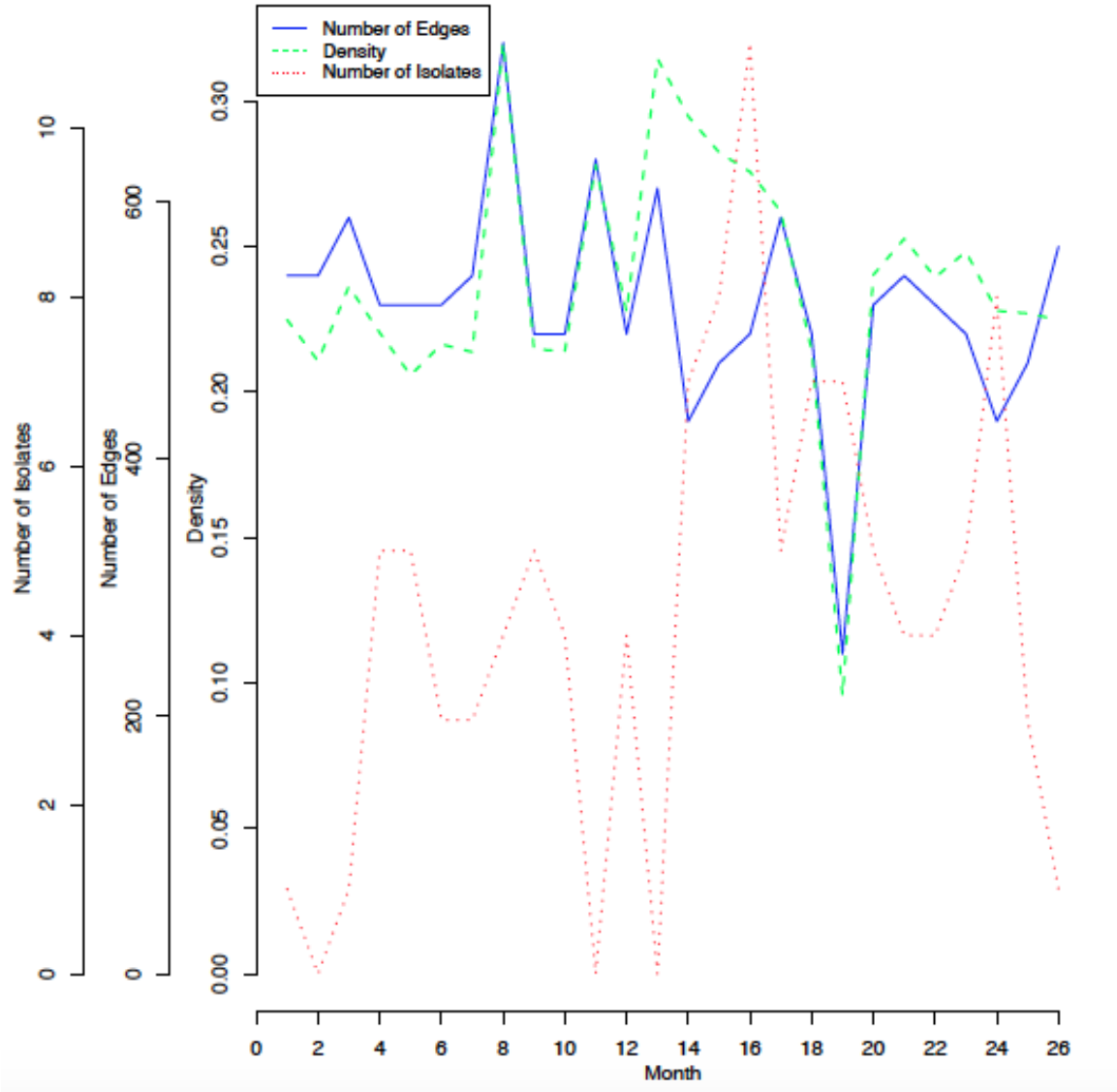

Figure 5.1. Change in network characteristics over time

Table 5.2. Network characteristics of online vs offline networks

Network	*Number of Nodes*	*Number of Edges*	*Density*	*Isolates*
Online	41	326	0.2	1
Advice	41	209	0.13	1
Brainstorming	41	174	0.11	1
Working	41	244	0.15	0

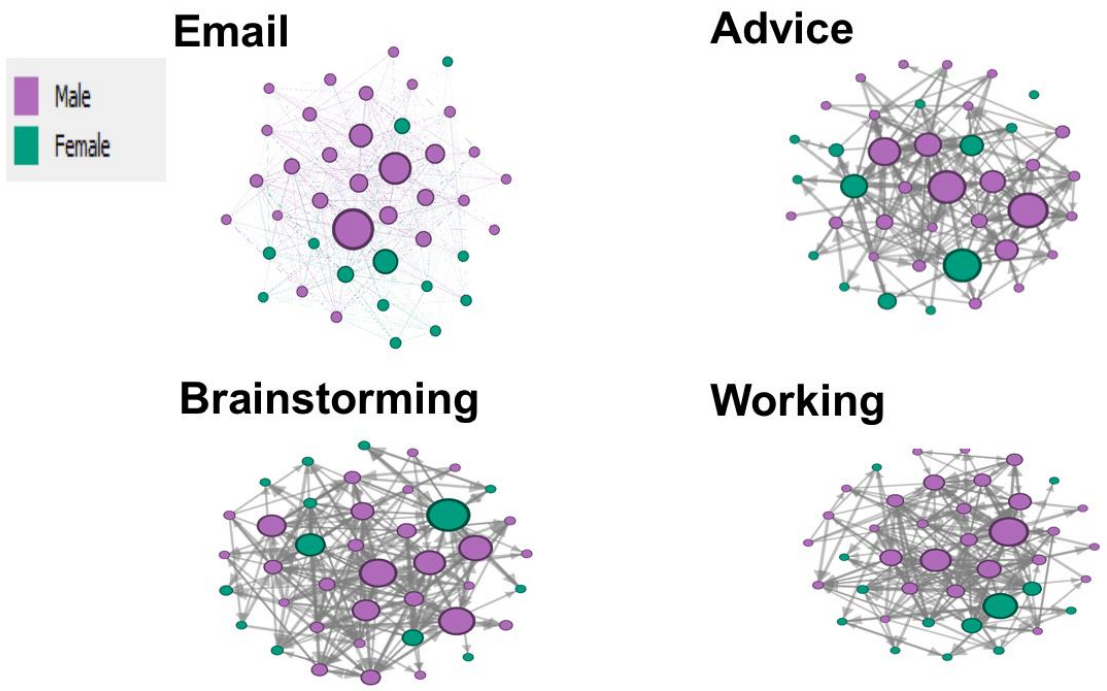

Figure 5.2. Plots of online vs offline networks

Factors Affecting the Formation of Online Social Networks

First, we fitted the 26 different networks corresponding to the different months, onto Model 1, to test hypothesis H1 (the tendency towards centralization). Social capital predicts a tendency towards centralization, and social comparison predicts a tendency towards decentralization. In all 26 months, the parameter for *gwidegree* was negative, indicating a tendency for low degree nodes in the networks, with only a few very high degree nodes. Negative values of *gwidegree* show that the network is centralized; hence, our findings lend support for the social capital perspective of online network formation. Fitting the overall network covering all email interactions between July 2015 and August 2017 inclusive onto Model 1 further confirmed this finding, hence we are confident that online networks evolve with a tendency towards centralization.

Second, we fitted the 26 networks onto Model 2, to test hypotheses H2 – H6. Positive values of *mutual* indicate tendency towards reciprocity, meaning that nodes tend to return incoming ties. In all months, the parameter for *mutual* was positive, indicating a tendency towards reciprocity. In this case, both social capital and social comparison predicted the same outcome, and they were both supported. Fitting the overall network onto Model 2 supported these results as well, further validating H2.

The greater the value of the *edges* parameter, the denser the network; hence a high value would support the social capital hypothesis and a low value the social comparison hypothesis. In all 26 networks, the coefficient for the *edges* term was negative, indicating a low tendency towards forming ties. However, the mean *edges* parameter was -3.26 (SD = 0.97) and is relatively high compared to other real world networks (Harris 2013). The *edges* parameter for the overall network was -1.51, further validating the finding that the network is becoming denser with time. This finding supports the social capital perspective of social network formation.

The effect of education on node popularity was inconsistent over time; highly educated nodes were more popular than lowly educated nodes in 8/26 months, less popular in 2/26, and just as popular in the remaining 16 months. High values of the education popularity parameter would support the social capital hypothesis and low values would support the social comparison hypothesis. Fitting the combined network onto Model 2 indicated that education did not have an effect on the popularity of individuals within the network. These findings lend do not support the social capital perspective of online network formation.

A positive value for the *nodematch* parameter indicates a tendency towards homophily. If social comparison was supported, we would detect gender homophily. In all 26 months, we detected no gender homophily in email ties using Model 2. Fitting the combined network onto Model 2 validated this finding. This shows that the social comparison perspective of online network evolution is not supported.

For H6, the effect of introversion/extraversion personality on gregariousness was inconsistent over the 26 months; in 2/26 months introverts were more gregarious, in 6/26 months extroverts were more so, and in the remaining 18 months there was no discernible effect. However, fitting the combined network onto Model 2 revealed homophily in tie formation according to personality type, which lends support towards the social comparison perspective. The results are detailed in Table 5.3 and summarized in Table 5.4.

The control variable, the number of known EmailCo contacts prior to joining the organization, had a positive relationship with popularity in the combined network. This means that individuals with many prior EmailCo contacts had more incoming ties than individuals with few contacts.

The above findings show that tie formation in online networks is a complex phenomenon that potentially requires multiple theoretical approaches to predict. They also show that cross-sectional studies might reveal biased findings; thus, when studying network formation, longitudinal analysis must be emphasized.

Differences Between Online and Offline Networks

To test H7, we fitted the July 2017 email network and the advice, brainstorming, and working networks (online and offline networks) onto Model 1. The *gwidegree*

parameter for the email network was more negative (-4.94) than that of the offline networks (-4.57, -3.69, and -4.38) for the advice, brainstorming, and working networks respectively. This finding suggests that the email network is much more centralized than the offline networks. This means that H7 is supported.

H8 was tested by fitting the 4 online and offline networks onto Model 3. The parameter for *mutual* is greater for the online network than for the offline networks, suggesting that reciprocity is greater in online networks. This finding does not support H8.

H9, tested using Model 3, is supported. The *edges* parameter for the online network is much higher than that of the offline networks (-3.57 vs -5.34, -5.88, and -4.28 for the advice, brainstorming, and working networks respectively).

H10 was tested using Model 3. There was no homophily recorded in each of the 4 networks, hence we did not find support for H10. However, demographic influences were more salient in offline networks than in the online network. The results are detailed in Table 5.5.

Results for the control variables are summarized in Table 5.6. Among the control variables, the importance afforded to face to face communication was positively related with online popularity. Further, employees working primarily at a non-headquarters location had more email network ties than employees that primarily work at headquarters.

In the advice network, task interdependence and email importance were positively related with gregariousness, phone importance and face to face importance were negatively related with gregariousness, and phone importance and face to face importance were positively related with popularity. Further, employees working

primarily at a non-headquarters location had fewer advice network ties than employees that primarily work at headquarters.

Task interdependence was positively associated with brainstorming network gregariousness i.e. individuals whose work highly depends on teamwork and coordination tend to brainstorm with more people than individuals with low teamwork and coordination requirements. Further, the number of prior EmailCo contacts was negatively associated with brainstorming popularity. Email importance was negatively associated with brainstorming gregariousness, and phone importance was negatively associated with brainstorming gregariousness. In the working network, task interdependence was positively associated with gregariousness, email importance was positively associated with gregariousness, and phone importance was negatively associated with gregariousness.

We used Model 4 to test H11. We found that individuals' positions within the online network are associated with their positions in the advice and working networks. Working popularity was negatively associated with online popularity, but working gregariousness was positively associated with online gregariousness. Advice gregariousness was negatively associated with online gregariousness. We found no relationship between online network position and brainstorming network position. These findings support H11. The findings are summarized in Table 5.7.

To compare ERGM models, we use the Akaike Information Criterion (AIC) (Akaike 1974; Harris 2013). The idea behind the AIC is that adding parameters usually results in a model that fits the data better than a simpler model.

Table 5.3. Results - Online network evolution.

Network	*Edges*	*GWID**	*Recip*	*H_Gen*	*Greg_Gen***	*Pop_Gen***	*H_Ed*	*Greg_Ed*	*Pop_Ed*	*H_Pers*	*Greg_Pers****	*Pop_Pers*
Jul-15	-4.11	-3.91	3.41				-0.25		0.26			
Aug-15	-4.05	-3.89	4.13					0.33				
Sep-15	-4.98	-5.02	3.28						0.17			-0.33
Oct-15	-4.46	-4.26	4.75				-0.30		0.29		0.37	0.38
Nov-15	-3.18	-4.92	3.98				-0.27					-0.2
Dec-15	-3.31	-4.19	3.78			0.45	-0.24					
Jan-16	-3.47	-4.66	4.32					0.16				
Feb-16	-2.59	-4.98	4.44			-0.34					0.29	-0.37
Mar-16	-2.16	-4.35	3.76		0.45							-0.33
Apr-16	-2.92	-4.79	3.64						-0.23			
May-16	-1.55	-6.12	3.17						0.14			-0.41
Jun-16	-3.13	-4.08	4.42		0.62	0.6		-0.33			0.52	-0.51
Jul-16		-5.13	3.58			-0.65					0.32	-0.65
Aug-16	-1.92	-5.18	3.96						-0.16		-0.33	-0.3
Sep-16	-3.18	-5.47	4.06						0.35			-0.35
Oct-16	-1.22	-5.14	3.80			-0.46			0.15		0.22	-0.52
Nov-16	-3.33	-5.24	3.71					0.18				
Dec-16	-2.69	-4.73	3.75									
Jan-17	-4.01	-4.43	3.71								-0.27	0.43
Feb-17	-4.37	-4.83	3.42					0.19				
Mar-17	-2.26	-6.28	3.28			0.37						
Apr-17	-4	-5.68	3.56									0.21
May-17	-3.47	-5.36	3.71			0.39			0.18			0.19
Jun-17	-3.41	-4.31	4.18				0.28				0.47	-0.42
Jul-17	-4.75	-4.56	3.18			0.56		0.25	0.18			
Aug-17	-2.95	-4.94	3.81		-0.45							0.22
Combined	-1.51	-8.02	5.04							0.20		-0.54

*Results from Model 1, the rest are from Model 2; **Reference group for gender is the female category; ***Personality score increases with level of extraversion; Recip – Reciprocity; GWID – Geometrically Weighted Indegree; H_Gen – Gender Homophily; Greg_Gen – Gender Gregariousness; Pop_Gen – Gender Popularity; H_Ed – Education Homophily; Greg_Ed – Education Gregariousness; Pop_Ed – Education Popularity; H_Pers – Personality Homophily ; Greg_Pers – Personality Gregariousness; Pop_Pers – Personality Popularity

Table 5.4. Results: Competing hypotheses according to social capital and social comparison theories

Hypothesis	*Social Capital Perspective*	*Social Comparison Perspective*	*Result – Supported Perspective*
H1	*Over time, an online network evolves with a tendency towards centralization.*	Over time, an online network evolves with a tendency towards decentralization.	Social capital
H2	*Over time, an online network evolves with a tendency towards reciprocity.*	*Over time, an online network evolves with a tendency towards reciprocity.*	Both
H3	*Over time, an online network evolves with a tendency towards high density of ties.*	Over time, an online network evolves with a tendency towards low density of ties.	Social capital
H4	Over time, higher educated nodes are more popular than lower educated nodes in online networks.		Not supported
H5		Over time, ties are more likely to form between people of the same gender than between people of different genders in online networks.	Not supported
H6		*Over time, ties are more likely to form between people of the same personality than between people of different personalities in online networks.*	Social comparison

The AIC formula penalizes more complex models (a model gains complexity with an increasing number of parameters). Given two competing models, the model with the smaller AIC is the better model. In the Appendix, we show detailed results of fitting our models onto all the networks. For each of the 26 networks over the period July 2015 to August 2017, we show detailed results from the Bernoulli model, Model 1 (Bernoulli +

centralization), and Model 2 (Bernoulli + reciprocity + demographic influences). In all months, Model 2 outperformed both Model 1 and the Bernoulli model, and Model 1 outperformed the Bernoulli model. Tables with the raw results from the analyses are found in Appendix B (Tables B1 – B92).

Post-hoc Analysis

Because Model 2 was outperformed by Model 1, we tested whether there are significant differences in the centralization scores of the 26 observed networks with randomly generated networks of similar size and density (Harris 2013). We tested the mean differences using one way analysis of variance. We found that the mean centralization score for the observed networks (mean = 0.51, sd = 0.15) was significantly larger than the mean for the randomly generated networks (mean = 0.16, sd = 0.03) at the 0.05 significance level ($F = 150.2$, $p = 0.000$, $df = 1$). These results lend further credence to the hypothesis that online networks demonstrate high centralization over time. For testing hypotheses 7 – 10, Model 3 outperformed both Model 1 and the Bernoulli model, and Model 1 outperformed the Bernoulli model.

Type I error – the possibility that we detected significant effects where none exist – was not a problem in this study. The odds ratios for the structural and demographic effects were high. For example, for the email network in August 2017, the parameter for reciprocity was 5.04. This translates to an odds ratio of 154 ($e^{5.04}$) which means a tie is 154 times more likely to form between two nodes if they already share a tie (Harris 2013). The smallest parameter in the combined July 2015 to August 2017 network was the parameter for personality homophily.

Table 5.5. Results - Differences between online and offline networks

Network	*Edges*	*GWID**	*Recip*	*Greg_Gen***	*Pop_Gen*	*Pop_Ed*	*Pop_Ed*	*Greg_Pers****	*Pop_Pers*
Email	-3.57	-4.94	3.93	-0.67					0.56
Advice	-5.34	-4.57	1.91	0.59	-0.42	0.13			-0.25
Brainstorming	-5.88	-3.69	1.96	0.67	-0.36				-0.19
Working	-4.96	-4.38	2.79	0.9	-0.74		0.21		-0.36

*Results from Model 1, the rest are from Model 2; **Reference group for gender is the female category; ***Personality score increases with level of extraversion; Recip – Reciprocity; GWID – Geometrically Weighted Indegree; Greg_Gen – Gender Gregariousness; Pop_Gen – Gender Popularity; Greg_Ed – Education Gregariousness; Pop_Ed – Education Popularity; Greg_Pers – Personality Gregariousness; Pop_Pers – Personality Popularity

Table 5.6. Correlations between network parameters and control variables.

Control Variable	*Popularity*	*Gregariousness*
Number of prior contacts	Brainstorming (-)*	
Task interdependence		Advice (+), Brainstorming (+), Working (+)
Email importance		Advice (+), Brainstorming (+), Working (+)
Face-to-face importance	Online (+), Advice (+)	Advice (-)
Phone importance	Advice (+)	Advice (-), Brainstorming (-), Working (-)

*A plus sign indicates a positive association between the control variable and the network parameter; a minus sign indicates a negative association.

Table 5.7. Results: Differences and interplay between online and offline networks

Hypothesis	*Description*	*Result*
H7	Centralization is greater in online networks than in offline networks.	Supported
H8	Reciprocity is greater in offline networks than in online networks.	Not supported
H9	Online networks are denser than offline networks.	Supported
H10	Homophily is greater in offline networks than in online networks.	Not supported
H11	The greater the synchronicity required in the activity underlying an offline network, the weaker the relationship between an individual's position within that network and his/her position in an online network.	Supported

The odds ratio for the effect was 1.22 (95% confidence interval = 1.18 – 1.26), meaning that a tie was 22% more likely between two nodes if they share similar personality (extrovert or introvert). These results demonstrate that the effect sizes of both structural and demographic variables were substantial and therefore provide evidence that our results do not reflect Type I error.

We also tested the effect of centrality of email network centrality on organizational commitment. We found that eigenvector centrality in the email network negatively predicts organizational commitment i.e. highly central members within the network expressed low levels of commitment to the organization. Hence, an individual's position within the email network is predictive of important organizational outcomes.

The next chapter discusses the results of our study and summarizes its research and practical implications.

CHAPTER SIX

Discussion

The social network perspective is a research approach that emphasizes the influence of relationships on individual, group, organizational, and societal outcomes. The effects of social networks on such varied outcomes as individual well-being, group performance, firm innovation etc. have been studied for decades, and it is likely that more effects will be revealed in the future. Yet, comparatively little research has investigated how these social networks form primarily because of (1) the difficulty of collecting social network data longitudinally and (2) methodological challenges in modeling the complex dependencies among the related sets of observations that characterize social networks. By tackling these two challenges, our study contributes towards the literature on organizational social networks in manifold ways.

We theorized how online network formation over a long period of time using two competing theories. Social capital theory predicts that individuals are motivated to form relationships in order to gain access to certain resources i.e. novel information, promotion opportunities, project opportunities etc. (Kilduff and Tsai 2003; Lin 1999). Social comparison, on the other hand, argues that individuals form relationships on the basis of similarity with others i.e. the more alike two individuals are, the greater their likelihood to form a tie (Festinger 1954; McPherson, Smith-Lovin, and Cook 2001). How might these two theories fare in explaining tie formation in online networks?

Using social capital, we hypothesized that an online network evolves with a tendency towards centralization; social comparison predicted de-centralization (H1). We found support for the social capital hypothesis. This finding suggests that over time, a few individuals become more and more influential in spreading information within the organization. Our analyses suggest that individuals become popular in the network for a variety of reasons e.g. their roles, personalities, levels of education etc. Current information builds upon, and is thus dependent on past information. It is thus more efficient for the network to (organically) designate a few individuals to store and maintain network wide information. Over time, these few individuals acquire an increasingly greater number of ties because of their current centrality within the network and this process repeats.

Both social capital and social comparison predict the emergence of reciprocity in online networks over time (H2). We found support for this hypothesis. Social capital theory suggests that reciprocity is an organizational norm (Lin 1999); hence, people reciprocate to others so as not to violate the norms that guide interpersonal interactions in organizations. Social comparison suggests that individuals compare the levels of effort exerted by their peers and respond accordingly; hence, when individuals receive email, they feel indebted to the email senders and respond in order to pay the debt.

While social capital theory predicted high online network density over time, social comparison predicted low density (social comparison posits that the network becomes segregated across demographic lines over time) (H3). We found support for the social capital hypothesis. This suggests that individuals seek to maximize social capital by making as many connections as possible. This finding was robust across even high

thresholds for tie inclusion. We initially defined a tie to exist if the median number of emails sent was exceeded, then raised the threshold to the 60th percentile and subsequently the third quartile, the network still remained dense, indicating that every individual still retains ties the height of the threshold notwithstanding.

The influences of demographic characteristics on network formation were each hypothesized differently by the social capital and social comparison theories. To hypothesize with the social capital perspective, one asks whether the demographic characteristic likely offers any differential value to individuals possessing that characteristic. Hence, because education endows one with expertise, social capital predicts that highly educated individuals will be rewarded with ties in the network. Social comparison, on the other hand, stresses the importance of similarity in determining tie formation. Hence, social comparison predicts that individuals sharing the same education level will form more online ties than is predicted by chance. Therefore, whereas social capital emphasizes the instrumental value of network ties, social comparison emphasizes the affective value of these ties. Further, social comparison emphasizes the egalitarian nature of the distribution of ties across categories (Buunk et al. 2000).

We found support no for the social comparison hypothesis regarding the influence of gender on online tie formation (H5). Social capital posits that gender similarity plays no role in influencing tie formation. Social comparison fared better regarding the influence of personality (homophily according to type) (H6). These findings suggest that demographic influences on network formation are less predictable than structural influences.

To summarize, our results show that social capital and social comparison theories are each better equipped at predicting only specific aspects of online network formation. While social capital is better at predicting the structural tendencies that emerge in an online network over time, social comparison does better at predicting demographic influences. Network centralization, for example, emerges because access to resources (information, promotion opportunities, time with top management etc.) is likely to be non-uniform, accruing disproportionately to a few people with the majority having relatively little access. Structural tendencies arise primarily because network ties are distributive mechanisms for resources within an organization. Because in social comparison tie formation is predicated on personal characteristics, it is ill-equipped to predict the structural characteristics of tie formation. On the other hand, social comparison stresses the affective nature of social relationships; hence, it is well-equipped to predict tie formation based on personal characteristics. Our findings show that online tie formation is a complex phenomenon. Individuals form online ties for both instrumental and affective purposes; hence, multiple theories might be required to predict tie formation in online networks.

Centrality in email networks is associated with certain positive outcomes e.g. job performance and client satisfaction (Gloor 2016); at the same time, it is also associated with certain negative outcomes e.g. reduced job satisfaction and increased perceived work overload. This is in contrast to traditional offline networks such as trust, friendship, and advice networks where centrality is generally associated with positive outcomes such as higher performance, increased job satisfaction, increased well-being, and improved status (Agneessens and Wittek 2008; Kane and Borgatti 2011). Our findings provide a

possible explanation for the bifurcated impact of email networks. We showed that high network centralization is a consistent feature of email networks. We argued using social capital theory that individuals come to occupy central positions within the network because of various reasons e.g. knowledge, technical expertise, and access to valuable information. Because high centralization characterizes email networks, individuals valued for their expertise and knowledge increasingly attract new ties, which burdens them to the extent that their levels of job satisfaction and organizational commitment are reduced; in other words, increasing social capital may prove a liability in online networks in the long run.

What's Different about Online Networks?[1]

Our study examined four different networks simultaneously within the organization – one online and three offline (advice, brainstorming, and working)[2]. Broadly, we hypothesized that centralization and density should be higher in online networks, and that reciprocity and homophily should be stronger in offline networks.

We found support for H7 i.e. the online network exhibits higher centralization than the offline networks. Online networks likely aggregate many functions e.g. an individual might use email to seek advice from and brainstorm and "work with" other individuals. Prominent individuals within the offline networks are more accessible using email rather than face to face interactions, as are individuals renowned for other qualities

---

[1] This question is posed as worded by Kane et al. (2012)

[2] Calling these networks offline might be a misnomer, but is consistent with prior research in social networks (see Kane et al. 2012). It is highly likely that some advice is given out through email, for example. It might be more appropriate for "offline" and "online" to refer to the methods by which data is collected, survey – offline and digital trace logs – online.

within the organization e.g. friendliness, being in the know, knowledge of company policies etc. According to social capital theory, individuals invest in relationships to obtain some payoff (Burt 2002; Lin 1999); the payoff for establishing ties with high value individuals in an organization is higher than connecting with a random network member. Email enables individuals to create and maintain relationships with such high value individuals across geographic and status boundaries (Herring 2008). Offline networks are less centralized than online networks because tie formation in the former may be limited by geographic and status boundaries. In other words, it is much harder for an individual to consistently seek advice from or brainstorm with someone else when not physically co-located.

Contrary to our hypothesis H8, we found higher levels of reciprocity in online networks, implying that it is easier to reciprocate in online networks relative to offline networks. This finding is explained by the limited effort it takes for individuals to maintain online ties relative to offline ties. Higher levels of reciprocity in organizational networks are associated with organizational stability and institutionalization, both of which contribute to the social capital base of an organization (Hanneman and Riddle 2005). Thus, one way of increasing an organization's overall social capital is by augmenting offline networks with online networks.

We found support for H9 i.e. online networks are denser than offline networks. Social capital theory explains this finding. The number of ties maintained by a human is limited by the size of a neocortex; maintaining a tie requires cognitive effort e.g. for remembering the state of the relationship, key details about one's connections (birthdays, anniversaries, weddings etc.), responding with feedback etc. (Dunbar 1992; Kudo and

Dunbar 2001). In online networks, one can delegate the cognitive effort for maintaining ties to technology; for example, the calendars that are embedded within many email tools can store reminders about key events, and email tools also have autoreply features that serve to maintain relationships. Thus, technology reduces the amount of cognitive effort required to maintain ties. Social capital posits that ties are assets; hence, humans will seek to maintain as many ties as possible. Using technology, individuals will thus maintain higher numbers of ties than they can with purely face to face interactions. The email network is resultantly much denser than the offline networks.

Our findings show that the offline networks had declining levels of similarity with the email network in this order: working, advice, and brainstorming. The email network was the densest, followed by the working, advice, and brainstorming networks respectively. The email network had the highest level of centralization; the advice and working networks had similar levels of centralization and were more centralized than the brainstorming network. The email network also exhibited the highest reciprocity, followed by the working network; the advice and brainstorming networks exhibited similar levels of reciprocity.

The positive relationship between the email and advice network centralities suggest that these two networks mirror each other, suggesting that email plays a significant role in disseminating advice within the organization. The negative relationship between the email and working network centralities means that highly influential individuals in the email network are less influential in the working network; this suggests that email may be proving a distraction from doing work within the organization. Email

and brainstorming network centralities are not related. This finding may be explained by the face to face nature of brainstorming work.

The method we used, exponential random graph modeling, is a simulation-based method in which a researcher suggests that the network arises out of structural and node characteristic influences. Through a series of simulations, a network emerges with parameters that maximize the probability of the observed network, given the size of the network. To answer RQ1 (what factors influence online network formation?), we tested our hypotheses in two different ways, first by fitting each of the networks corresponding to the 26 different months between July 2015 and August 2017 (the period over which we have email logs) onto two ERGM models Model 1 (Bernoulli model + centralization) and Model 2 (Bernoulli model + centralization + demographic influences), and second by fitting the combined network onto Model 1 and Model 2. The findings about structural tendencies were consistent across time and across the two methods, lending high confidence in our findings. The findings on demographic influences were less straightforward using each method in isolation, but by combining insights from both sets of results, we were able to build a clearer picture of the ways in which demographic characteristics influence online network formation.

Implications

Research

Structural influences are stable over time; thus, it is valid to examine them cross-sectionally. Demographic influences, on the other hand, are fickle; in some months certain influences were significant, in others they were not. This suggests that findings from research conducted on cross-sectional snapshots of social networks might be biased.

The implication is that work seeking to establish how demographic characteristics influence network formation must necessarily be longitudinal.

A second implication for research concerns methodology. Thus far, some research on social networks has depended on experiments; for example, studies have created artificial social networks to understand the effects of prior social ties on economic exchange (Bapna et al. 2011; Suri and Watts 2011). With online networks, the limitations of such experiments are obviated. Researchers could examine the effects of prior social ties on purchase decisions on platforms such as Craigslist and E-bay. To test whether social comparison generalizes to online networks, future research could examine whether the length of a response email is related to the length of the original email. Social comparison posits that individuals respond to the amount of effort exerted by their connections; hence, a positive relationship between email length and response length would help validate the generalizability of the theory. These examples show that online networks can ease some of the difficulties that characterize existing social network research.

It is commonly stated in the research that networks change over time, but how they change is relatively unexplored (see Borgatti et al. 2009; Yan, Peng, and Tan 2015). We have provided evidence that some aspects of online network evolution are predictable. For example, the structural attributes of centralization and reciprocity are relatively constant, but density fluctuates. On the other hand, the effect of node characteristics is quite unpredictable; for example, at times homophily according to education or gender may exist, other times it does not. Future research could explore the

conditions under which homophily disappears, perhaps as a result of changing network density, for example.

Practice

We have shown that online networks and offline networks resemble each other in some respects. Past research shows that centrality in offline networks predicts multiple outcomes of interest. However, the process for obtaining offline network relationship data is tedious and subject to many biases e.g. inaccurate recall and social desirability (Kane et al. 2012). The process is much more complex when examining network evolution; one needs to repeatedly survey organization members over a long period of time, which will likely fatigue them. Instead of repeatedly surveying organization members to monitor network change, practitioners can monitor their email networks and other social media platforms to understand how their networks change over time.

Centrality in email networks is associated with decreased job satisfaction (Merten and Gloor 2010). Examining email network formation reveals the demographic profiles of highly central individuals within the network. Such information reveals whether any systematic demographic differences in centrality exist. In several months in EmailCo's network, education level was positively associated with popularity; hence, highly educated individuals within the company would have been at higher risk of decreased job satisfaction in those months. It is possible that the most central individuals might be those valued for their skills and knowledge within the organization, in which case the attendant decreased job satisfaction might motivate them to seek opportunities elsewhere, to the detriment of the organization.

Indeed, we showed in the post-hoc analysis that email network centrality negatively predicts an individual's level of organizational commitment. A possible explanation for this finding is that highly central members within the email network are overloaded with email, to the extent that it gives them negative work experiences, which in turn likely reduces organizational commitment. As such, if practitioners monitor their networks, it becomes possible to identify employees that disproportionately receive emails from others and possibly intervene to reduce such email workload.

Limitations

The findings from our study must be considered within the context of three main limitations. First, communication in any organization is not likely to be limited to email. Some official communication is conducted through face to face interactions, phone conversations, texts, chat etc. As such, the email communication network might distort the true nature of communication relationships in the organization. We controlled for the level of importance assigned to email, face-to-face, and phone interactions, but it is not possible to control for every conceivable means of communication. Nevertheless, email remains the most important form of communication within organizations (Gloor 2016).

Second, because we do not have access to actual email message content, we treat all emails the same. In real organizational settings, however, an individual is likely to vary his/her response to an email based on the email's perceived importance. Future research should examine whether content type influences email network evolution; for example, do people that primarily send instructional content (instructions or directives) tend to be more central than people that do not. If central members within the email

network generally receive directives, it might explain why they would express lower levels of organizational commitment.

Third, numerous other factors might also predict network formation. Our study only considered the introversion/extraversion aspect of personality, but did not consider conscientiousness, neuroticism, openness, and agreeableness as possible factors affecting how individuals form relationships (Hayes and Joseph 2003). Future research should examine the influences of these factors on network formation, to build a clearer picture. Other possible factors for possible future investigation include organizational commitment, job satisfaction, and well-being. Examining these factors may help establish whether what are commonly perceived as outcomes of network position are in fact antecedents of network formation.

CHAPTER SEVEN

Conclusion

In this study, we investigated the factors that influence online social network formation. To uncover these factors, we tracked changes in the composition of the network of email communications among employees of a technology company over a period of 26 months. We found that the structural influences of high centralization, high density, and high reciprocity characterize online networks as they change over time. On the other hand, demographic influences such as popularity according to education and homophily according to gender were not consistent over time. As such, studies that examine the demographic factors affecting network formation must necessarily be longitudinal rather than cross-sectional, as has been the case to date.

Further, we examined the structural differences between online and offline social networks. We found that online networks exhibit higher density, higher reciprocity, and higher centralization than offline networks. These findings suggest that online networks are fundamentally structurally different from offline networks. Because online networks have high levels of density, centralization, and reciprocity, adopting electronic tools of communication offers organizations with means of building their combined social capitals, beyond the usual advantages of speed, affordability, and convenience of using these tools.

Last, we examined how online networks interact with offline networks within the organization. In answer to the question that has been posed in previous research,

concerning how online networks interact with offline networks, we found that the nature of these interactions partly depend on the level of synchronicity required in the activity that underlies the offline networks. Brainstorming between any two individuals requires higher synchronicity than either working together or sharing advice. Thus, electronic communication might be ill-suited for brainstorming. Indeed, we found significant associations between email network positions and working and advice network positions but none with brainstorming network positions. These findings imply that in addition to exploring the differences between offline and online networks, social network researchers should also identify dimensions on which offline networks vary, for a more nuanced understanding of these differences.

APPENDICES

APPENDIX A

Questions

Survey Questions

The tables below show the survey questions for determining levels of extraversion, task interdependence, and organizational commitment and the associated mean response scores in EmailCo. Chronbach's alpha exceeded 0.7 for all three constructs; 0.7 is the benchmark that indicates that a construct has high internal consistency (Santos 1999).

Table A.1. Introversion/extraversion (Sato 2005)

Question	*Mean*	*SD*
1. Are you a talkative person?	3.18	1.00
2. Are you rather lively?	3.25	0.99
3. Do you enjoy meeting new people?	3.63	0.94
4. Can you usually let yourself go and enjoy yourself at a lively party?	3.40	0.99
5. Do you usually take the initiative in making new friends?	3.30	1.00
6. Can you easily get some life into a rather dull party?	2.95	0.95
7. Do you tend to keep in the background on social occasions?	2.48	1.07
8. Do you like mixing with people?	3.48	0.89
9. Do you like plenty of action and excitement around you?	3.53	0.81
10. Do other people think of you as being very lively?	3.13	1.10
11. Can you get a party going?	2.78	1.15
12. Are you mostly quiet when you are with other people?	2.05	0.86

Cronbach's alpha = 0.93

Table A.2. Task interdependence (Langfred 2005)

Question	*Mean*	*SD*
1. The team works best when we coordinate our work closely.	5.98	1.17
2. Team members have to work together to get group tasks done.	5.75	1.37
3. The way individual members perform their jobs has a significant impact on others in the team.	6.20	1.14
4. My work cannot be done unless other people do their work.	5.30	1.35
5. Most of my work activities are affected by the activities of other people on the team.	5.08	1.37
6. Team members frequently have to coordinate their efforts with each other.	5.73	1.00
7. We cannot complete a project unless everyone contributes.	5.43	1.66

Cronbach's alpha = 0.78

Table A.3. Organizational commitment (Tsui et al. 1997)

Question	*Mean*	*SD*
1. I am willing to put in effort beyond the norm for the success of the organization.	6.47	0.73
2. For me, this is the best of all possible organizations for which to work.	4.89	1.36
3. I am extremely glad to have chosen this organization to work for over other organizations.	5.61	1.13
4. This organization inspires the very best in the way of job performance.	5.18	1.28
5. I show by my actions that I really care about the fate of this organization.	6.08	0.88

Cronbach's alpha = 0.83

Network Questions

1. Advice network: Indicate the extent to which you turn to each of the following people for expert advice about work-related activities (Cross, Borgatti, and Parker 2001).
2. Brainstorming network: Indicate the extent to which you innovate or brainstorm together with the following people.
3. Working network: Indicate the extent to which you work together with the following people.

APPENDIX B

Interpreting ERGM Parameters

To demonstrate how to interpret the ERGM parameters, we use results from July 2015 (Table B.3). The baseline probability for tie formation is calculated from the *edges* parameter (Goodreau 2007; Harris 2013; Hunter et al. 2008). This means that, in the Bernoulli model, the conditional log-odds of a tie between any two people equals the *edges* parameter. Increasing variables to the Bernoulli model results in a decrease of the *edges* parameter when tie formation is not random (true for most real-world networks). In a model with other variables, the *edges* parameter equals the baseline conditional log-odds of a tie between two people:

$$\text{Probability of tie formation} = \frac{e^{\text{estimate}}}{1+e^{\text{estimate}}} = \frac{e^{-4.11}}{1+e^{-4.11}} = 0.0161$$

Hence, any pair of people A and B have a 1.6% chance of forming a tie in the email network of July 2015. However, if one tie already exists between A and B, the probability that they will form a tie in the opposite direction rises, because the *mutual* parameter is positive:

$$\begin{aligned}\text{Conditional logodds of reciprocal tie} &= \text{edges} + \text{mutual}\\ &= -4.11 + 3.41\\ &= -0.7\end{aligned}$$

Hence, for two people with a tie existing in one direction, the probability that a tie forms in the other direction is given by:

$$\text{Probability of tie formation} = \frac{e^{\text{estimate}}}{1+e^{\text{estimate}}} \qquad (2)$$
$$= \frac{e^{-0.7}}{1+e^{-0.7}}$$
$$= 0.332$$

Hence, the probability that two individuals form a tie rises to 33.2% if a tie already exists between them.

Suppose that A and B knew 3 and 7 people respectively prior to joining EmailCo. The conditional log-odds of forming a tie between them is now given by:

$$\text{Conditional logodds of incoming tie}$$
$$= \text{edges} + \text{mutual} + \text{nodeicov.Prior\_Contacts} * (\text{Prior\_Contacts(A)} + \text{Prior\_Contacts(B)})$$
$$= -4.11 + 3.41 - 0.16(3+7)$$
$$= -2.3$$

$$Probability\ of\ tie\ formation = \frac{e^{estimate}}{1+e^{estimate}}$$
$$= \frac{e^{-2.3}}{1+e^{-2.3}}$$
$$= 0.091$$

Thus, the probability of tie formation decreases to 9% in this case, because the greater the number of prior contacts, the lower the probability of tie formation.

Tables B1 to B92 show the results of fitting the networks onto the Bernoulli model, Model 1, Model 2, Model 3, and Model 4.

Model 1 is mathematically expressed as follows, with κ representing the sum of the weighted statistics across all possible networks of the same size as the observed networks, and θ_k representing the weight of each statistic.

$$\text{Prob(Network)} = \frac{1}{\kappa}\exp(\text{Edges} + \theta_1 \text{Centralization})$$

Model 2 is expressed in similar form:

$$\text{Prob(Network)} = \frac{1}{\kappa}\exp(\text{Edges} + \theta_1 \text{Reciprocity} + \theta_2 \text{Homophily(Gender)} + \theta_3 \text{Popularity (Gender)} + \theta_4 \text{Gregariousness (Gender)} + \theta_5 \text{Homophily(Education)} + \theta_6 \text{Popularity (Education)} + \theta_7 \text{Gregariousness (Education)} + \theta_8 \text{Homophily(Personality)} + \theta_9 \text{Popularity (Personality)} + \theta_{10} \text{Gregariousness (Personality)} + \theta_{11} \text{Popularity (Prior\_Contacts)} + \theta_{12} \text{Gregariousness (Prior Contacts)})$$

Model 3:

$$\text{Prob(Network)} = \frac{1}{\kappa}\exp(\text{Edges} + \theta_1 \text{Reciprocity} + \theta_2 \text{Homophily(Gender)} + \theta_3 \text{Popularity (Gender)} + \theta_4 \text{Gregariousness (Gender)} + \theta_5 \text{Homophily(Education)} + \theta_6 \text{Popularity (Education)} + \theta_7 \text{Gregariousness (Education)} + \theta_8 \text{Homophily(Personality)} + \theta_9 \text{Popularity (Personality)} + \theta_{10} \text{Gregariousness (Personality)} + \theta_{11} \text{Popularity (Prior\_Contacts)} + \theta_{12} \text{Gregariousness (Prior Contacts)} + \theta_{13} \text{Popularity (Email\_Importance)} + \theta_{14} \text{Gregariousness (Email\_Importance)} + \theta_{15} \text{Popularity(F2F\_Importance)} + \theta_{16} \text{Gregariousness(F2F\_Importance)} + \theta_{17} \text{Popularity(Phone\_Importance)} + \theta_{18} \text{Gregariousness(Phone\_Importance)} + \theta_{19} \text{Popularity(Task\_Interdependence)} + \theta_{20} \text{Gregariousness(Task\_Interdependence)})$$

Model 4:

$$\text{Prob(Network)} = \frac{1}{\kappa}\exp(\text{Edges} + \theta_1 \text{Reciprocity} + \theta_2 \text{Homophily(Gender)} + \theta_3 \text{Popularity (Gender)} + \theta_4 \text{Gregariousness (Gender)} + \theta_5 \text{Homophily(Education)} + \theta_6 \text{Popularity (Education)} + \theta_7 \text{Gregariousness (Education)} + \theta_8 \text{Homophily(Personality)} + \theta_9 \text{Popularity (Personality)} + \theta_{10} \text{Gregariousness (Personality)} + \theta_{11} \text{Popularity (Prior\_Contacts)} + \theta_{12} \text{Gregariousness (Prior\_Contacts)} + \theta_{12} \text{Popularity (Advice\_Centrality)} + \theta_{13} \text{Gregariousness (Advice\_Centrality)} + \theta_{14} \text{Popularity (Brainstorming\_Centrality)} + \theta_{15} \text{Gregariousness(Brainstorming\_Centrality)} + \theta_{16} \text{Popularity (Work\_Centrality)} + \theta_{17} \text{Gregariousness (Work\_Centrality)} + \theta_{18} \text{Popularity (Prior\_Contacts)} + \theta_{19} \text{Gregariousness (Prior\_Contacts)} + \theta_{20} \text{Popularity (Email\_Importance)} + \theta_{21} \text{Gregariousness (Email\_Importance)} + \theta_{22} \text{Popularity(F2F\_Importance)} + \theta_{23} \text{Gregariousness(F2F\_Importance)} + \theta_{24} \text{Popularity(Phone\_Importance)} + \theta_{25} \text{Gregariousness(Phone\_Importance)} + \theta_{26} \text{Popularity(Task\_Interdependence)} + \theta_{27} \text{Gregariousness(Task\_Interdependence)})$$

Tables of Results

Table B.1. Bernoulli model results for EmailCo network (July 2015).

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.4683	0.0633	0	0.000000***

AIC: 1583 BIC: 1589

Table B.2. Model 1 results for EmailCo email network (July 2015).

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.05082	0.06159	0	0.000000***
gwidegree	-3.91207	0.2849	0	0.000000***

AIC: 1485 BIC: 1496

Table B.3. Model 2 results for EmailCo email network (July 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
Edges	-4.112455	0.480064	0	0.000000***
Mutual	3.408322	0.240626	0	0.000000***
nodematch.Gender	0.056918	0.126155	0	0.651922
nodeofactor.Gender.Male	0.014123	0.19445	0	0.942111
nodeifactor.Gender.Male	0.186324	0.192259	0	0.332623
nodematch.Education_Dich	-0.25413	0.136509	0	0.062835.
nodeocov.Education	0.002008	0.091286	0	0.982451
nodeicov.Education	0.256534	0.09176	0	0.005239**
nodematch.Personality_Dich	0.072005	0.103698	0	0.487545
nodeocov.Personality	0.115264	0.116939	0	0.324438
nodeicov.Personality	0.042181	0.116636	0	0.717661
nodeocov.Prior_Contacts	-0.04179	0.037823	0	0.269375
nodeicov.Prior_Contacts	-0.163954	0.04669	0	0.000458***

AIC: 1286 BIC: 1357

Table B.4. Bernoulli model results for EmailCo network (August 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.52131	0.06438	0	0.000000***

AIC: 1545 BIC: 1550

Table B.5. Model 1 results for EmailCo email network (August 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.06786	0.06085	0	0.000000***
gwidegree	-3.89458	0.28406	0	0.000000***

AIC: 1441 BIC: 1452

Table B.6. Model 2 results for EmailCo email network (August 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-4.047294	0.544188	0	0.000000***
mutual	4.132912	0.271266	0	0.000000***
nodematch.Gender	0.035192	0.123719	0	0.7761
nodeofactor.Gender.Male	-0.020117	0.212824	0	0.92471
nodeifactor.Gender.Male	0.083085	0.215874	0	0.70038
nodematch.Education_Dich	-0.123465	0.149771	0	0.40986
nodeocov.Education	0.331789	0.111977	0	0.00309**
nodeicov.Education	0.183404	0.111881	0	0.10135
nodematch.Personality_Dich	0.029427	0.100031	0	0.76866
nodeocov.Personality	-0.104111	0.132677	0	0.43275
nodeicov.Personality	-0.325406	0.133526	0	0.01491*
nodeocov.Prior_Contacts	-0.004571	0.033165	0	0.8904
nodeicov.Prior_Contacts	0.023714	0.032976	0	0.47216

AIC: 1175 BIC: 1245

Table B.7. Bernoulli model results for EmailCo network (September 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.1715	0.0581	0	0.000000***

AIC: 1797 BIC: 1802

Table B.8. Model 1 results for EmailCo email network (September 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.78678	0.05819	0	0.000000***
gwidegree	-5.02082	0.30518	0	0.000000***

AIC: 1642 BIC: 1653

Table B.9. Model 2 results for EmailCo email network (September 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-4.978553	0.459636	0	0.000000***
mutual	3.277538	0.223204	0	0.000000***
nodematch.Gender	0.034196	0.118477	0	0.772903
nodeofactor.Gender.Male	-0.069105	0.180165	0	0.701351
nodeifactor.Gender.Male	0.213688	0.172192	0	0.214788
nodematch.Education_Dich	-0.095576	0.128451	0	0.456942
nodeocov.Education	0.079969	0.085637	0	0.350534
nodeicov.Education	0.169416	0.087097	0	0.051930.
nodematch.Personality_Dich	0.065651	0.095083	0	0.490003
nodeocov.Personality	0.019512	0.111178	0	0.860707
nodeicov.Personality	0.388952	0.110099	0	0.000423***
nodeocov.Prior_Contacts	-0.003507	0.028462	0	0.901945
nodeicov.Prior_Contacts	-0.042207	0.030319	0	0.164091

AIC: 1470 BIC: 1540

Table B.10. Bernoulli model results for EmailCo email network (October 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.50893	0.06412	0	0.000000***

AIC: 1554 BIC: 1559

Table B.11. Model 1 results for EmailCo email network (October 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.99575	0.06058	0	0.000000***
gwidegree	-4.25881	0.28259	0	0.000000***

AIC: 1415 BIC: 1426

Table B.12. Model 2 results for EmailCo email network (October 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-4.46255	0.50336	0	0.000000***
mutual	4.75388	0.30466	0	0.000000***
nodematch.Gender	-0.04329	0.11666	0	0.7107
nodeofactor.Gender.Male	-0.12683	0.22642	0	0.5755
nodeifactor.Gender.Male	-0.21523	0.22729	0	0.3438
nodematch.Education_Dich	-0.30156	0.14384	0	0.0362*
nodeocov.Education	0.03703	0.11125	0	0.7393
nodeicov.Education	0.28911	0.11333	0	0.0108*
nodematch.Personality_Dich	0.03502	0.09638	0	0.7164
nodeocov.Personality	0.36744	0.14785	0	0.0130*
nodeicov.Personality	-0.37457	0.14711	0	0.0110*
nodeocov.Prior_Contacts	-0.01104	0.03342	0	0.7412
nodeicov.Prior_Contacts	0.06943	0.03215	0	0.0309*

AIC: 1143 BIC: 1214

Table B.13. Bernoulli model results for EmailCo network (November 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.24728	0.05931	0	0.000000***

AIC: 1743 BIC: 1749

Table B.14. Model 1 results for EmailCo email network (November 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.8212	0.0568	0	0.000000***
gwidegree	-4.9182	0.2949	0	0.000000***

AIC: 1584 BIC: 1595

Table B.15. Model 2 results for EmailCo email network (November 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-3.179471	0.443364	0	0.000000***
mutual	3.979754	0.240988	0	0.000000***
nodematch.Gender	-0.030528	0.104592	0	0.7704
nodeofactor.Gender.Male	0.031848	0.186556	0	0.8645
nodeifactor.Gender.Male	-0.25769	0.189805	0	0.1748
nodematch.Education_Dich	-0.269502	0.13168	0	0.0409*
nodeocov.Education	0.139152	0.095762	0	0.1464
nodeicov.Education	0.140852	0.095157	0	0.139
nodematch.Personality_Dich	-0.011598	0.093069	0	0.9008
nodeocov.Personality	0.027268	0.120221	0	0.8206
nodeicov.Personality	-0.203883	0.12026	0	0.0902.
nodeocov.Prior_Contacts	0.002181	0.030332	0	0.9427
nodeicov.Prior_Contacts	0.03138	0.029095	0	0.2809

AIC: 1374 BIC: 1444

Table B.16. Bernoulli model results for EmailCo network (December 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.50893	0.06412	0	0.000000***

AIC: 1554 BIC: 1559

Table B.17. Model 1 results for EmailCo email network (December 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.00888	0.06359	0	0.000000***
gwidegree	-4.18646	0.28596	0	0.000000***

AIC: 1425 BIC: 1436

Table B.18. Model 2 results for EmailCo email network (December 2015)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-3.313447	0.458859	0	0.000000***
mutual	3.778272	0.256306	0	0.000000***
nodematch.Gender	-0.005675	0.131958	0	0.9657
nodeofactor.Gender.Male	0.17388	0.212034	0	0.4123
nodeifactor.Gender.Male	0.454822	0.209477	0	0.0301*
nodematch.Education_Dich	-0.235761	0.130931	0	0.0719.
nodeocov.Education	0.03298	0.093908	0	0.7255
nodeicov.Education	-0.052363	0.093326	0	0.5748
nodematch.Personality_Dich	-0.019559	0.102116	0	0.8481
nodeocov.Personality	-0.068335	0.125743	0	0.5869
nodeicov.Personality	0.097045	0.125063	0	0.4379
nodeocov.Prior_Contacts	0.065668	0.028981	0	0.0236*
nodeicov.Prior_Contacts	0.052001	0.028836	0	0.0715.

AIC: 1231 BIC: 1301

Table B.19. Bernoulli model results for EmailCo network (January 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.24377	0.05925	0	0.000000***

AIC: 1746 BIC: 1751

Table B.20. Model 1 results for EmailCo email network (January 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.86309	0.05696	0	0.000000***
gwidegree	-4.66339	0.29565	0	0.000000***

AIC: 1613 BIC: 1624

Table B.21. Model 2 results for EmailCo email network (January 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-3.466209	0.408328	0	0.000000***
mutual	4.31519	0.261591	0	0.000000***
nodematch.Gender	-0.007008	0.099353	0	0.9438
nodeofactor.Gender.Male	0.165337	0.203533	0	0.4167
nodeifactor.Gender.Male	-0.228109	0.206149	0	0.2687
nodematch.Education_Dich	0.014232	0.119858	0	0.9055
nodeocov.Education	0.164303	0.096144	0	0.0877.
nodeicov.Education	-0.068952	0.093497	0	0.4609
nodematch.Personality_Dich	-0.084966	0.092648	0	0.3592
nodeocov.Personality	0.066761	0.126477	0	0.5977
nodeicov.Personality	0.003896	0.124834	0	0.9751
nodeocov.Prior_Contacts	-0.041609	0.032865	0	0.2057
nodeicov.Prior_Contacts	0.032446	0.030905	0	0.2939

AIC: 1345 BIC: 1415

Table B.22. Bernoulli model results for EmailCo network (February 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.23676	0.05913	0	0.000000***

AIC: 1751 BIC: 1756

Table B.23. Model 1 results for EmailCo email network (February 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.86309	0.05696	0	0.000000***
gwidegree	-4.66339	0.29565	0	0.000000***

AIC: 1590 BIC: 1601

Table B.24. Model 2 results for EmailCo email network (February 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-2.59117	0.41526	0	0.000000***
mutual	4.43753	0.2756	0	0.000000***
nodematch.Gender	-0.07619	0.09889	0	0.44115
nodeofactor.Gender.Male	0.25395	0.20315	0	0.21144
nodeifactor.Gender.Male	-0.33508	0.20026	0	0.09447.
nodematch.Education_Dich	-0.16418	0.12092	0	0.17473
nodeocov.Education	-0.03555	0.09598	0	0.71111
nodeicov.Education	0.03834	0.0943	0	0.68438
nodematch.Personality_Dich	-0.05024	0.09607	0	0.6011
nodeocov.Personality	0.29436	0.13299	0	0.02701*
nodeicov.Personality	-0.37374	0.13082	0	0.00433**
nodeocov.Prior_Contacts	0.02678	0.0306	0	0.38171
nodeicov.Prior_Contacts	0.04832	0.03065	0	0.11516

AIC: 1323 BIC: 1393

Table B.25. Bernoulli model results for EmailCo network (March 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.6021	0.0661	0	0.000000***

AIC: 1485 BIC: 1491

Table B.26. Model 1 results for EmailCo email network (March 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.99903	0.06261	0	0.000000***
gwidegree	-4.35451	0.28497	0	0.000000***

AIC: 1327 BIC: 1337

Table B.27. Model 2 results for EmailCo email network (March 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-2.16201	0.47126	0	0.000000***
mutual	3.75789	0.25911	0	0.000000***
nodematch.Gender	-0.21436	0.15231	0	0.1595
nodeofactor.Gender.Male	0.44705	0.21998	0	0.04229*
nodeifactor.Gender.Male	0.1876	0.22066	0	0.39534
nodematch.Education_Dich	0.06047	0.12448	0	0.62718
nodeocov.Education	0.04204	0.09353	0	0.6531
nodeicov.Education	-0.09416	0.0921	0	0.30675
nodematch.Personality_Dich	-0.01046	0.11097	0	0.92488
nodeocov.Personality	0.12279	0.12475	0	0.32513
nodeicov.Personality	-0.33086	0.12177	0	0.00665**
nodeocov.Prior_Contacts	-0.06716	0.03604	0	0.06256.
nodeicov.Prior_Contacts	0.02052	0.03252	0	0.52818

AIC: 1210 BIC: 1280

Table B.28. Bernoulli model results for EmailCo email network (April 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.18845	0.05837	0	0.000000***

AIC: 1785 BIC: 1790

Table B.29. Model 1 results for EmailCo email network (April 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.82811	0.05709	0	0.000000***
gwidegree	-4.78714	0.30342	0	0.000000***

AIC: 1648 BIC: 1659

Table B.30. Model 2 results for EmailCo email network (April 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-2.923641	0.390865	0	0.000000***
mutual	3.638926	0.237216	0	0.000000***
nodematch.Gender	-0.072499	0.114935	0	0.52827
nodeofactor.Gender.Male	0.041245	0.186417	0	0.82492
nodeifactor.Gender.Male	0.387624	0.188691	0	0.04011*
nodematch.Education_Dich	0.009547	0.112811	0	0.93257
nodeocov.Education	0.083553	0.082074	0	0.30882
nodeicov.Education	-0.232148	0.081766	0	0.00458**
nodematch.Personality_Dich	-0.027254	0.094768	0	0.7737
nodeocov.Personality	0.130101	0.112148	0	0.24619
nodeicov.Personality	0.166218	0.112808	0	0.14082
nodeocov.Prior_Contacts	-0.099497	0.036043	0	0.00584**
nodeicov.Prior_Contacts	0.007512	0.032356	0	0.81645

AIC: 1438 BIC: 1508

Table B.31. Bernoulli model results for EmailCo network (May 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.8213	0.05361	0	0.000000***

AIC: 2021 BIC: 2026

Table B.32. Model 1 results for EmailCo email network (May 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.53804	0.05861	0	0.000000***
gwidegree	-6.12421	0.34335	0	0.000000***

AIC: 1859 BIC: 1870

Table B.33. Model 2 results for EmailCo email network (May 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.5527	0.40475	0	0.00013***
mutual	3.1656	0.20432	0	0.000000***
nodematch.Gender	-0.06598	0.10336	0	0.5233
nodeofactor.Gender.Male	-0.17174	0.1598	0	0.28268
nodeifactor.Gender.Male	0.07123	0.15812	0	0.65245
nodematch.Education_Dich	-0.07242	0.11069	0	0.51307
nodeocov.Education	0.02589	0.07569	0	0.73232
nodeicov.Education	0.1402	0.07552	0	0.06356.
nodematch.Personality_Dich	-0.05033	0.08557	0	0.55652
nodeocov.Personality	0.09745	0.0994	0	0.32707
nodeicov.Personality	-0.41499	0.09769	0	0.000000***
nodeocov.Prior_Contacts	0.02023	0.02504	0	0.41909
nodeicov.Prior_Contacts	-0.06716	0.02608	0	0.01010*

AIC: 1697 BIC: 1767

Table B.34. Bernoulli model results for EmailCo email network (June 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.6646	0.0675	0	0.000000***

AIC: 1439 BIC: 1445

Table B.35. Model 1 results for EmailCo email network (June 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.06745	0.06295	0	0.000000***
gwidegree	-4.08175	0.28426	0	0.000000***

AIC: 1304 BIC: 1315

Table B.36. Model 2 results for EmailCo email network (June 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-3.12802	0.46772	0	0.000000***
mutual	4.41561	0.30726	0	0.000000***
nodematch.Gender	-0.274	0.18635	0	0.141675
nodeofactor.Gender.Male	0.6246	0.26486	0	0.018482*
nodeifactor.Gender.Male	0.60162	0.25975	0	0.020675*
nodematch.Education_Dich	0.08929	0.13128	0	0.496507
nodeocov.Education	-0.32794	0.09914	0	0.000961***
nodeicov.Education	0.15238	0.09969	0	0.126573
nodematch.Personality_Dich	-0.05286	0.10589	0	0.617675
nodeocov.Personality	0.51501	0.13859	0	0.000209***
nodeicov.Personality	-0.50643	0.13646	0	0.000213***
nodeocov.Prior_Contacts	0.02256	0.03705	0	0.542704
nodeicov.Prior_Contacts	-0.02139	0.03777	0	0.571278

AIC: 1097 BIC: 1167

Table B.37. Bernoulli model results for EmailCo email network (July 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.0536	0.0564	0	0.000000***

AIC: 1877 BIC: 1882

Table B.38. Model 1 results for EmailCo email network (July 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.74121	0.05464	0	0.000000***
gwidegree	-5.13263	0.30694	0	0.000000***

AIC: 1741 BIC: 1752

Table B.39. Model 2 results for EmailCo email network (July 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.271379	0.429719	0	0.527785
mutual	3.583616	0.243264	0	0.000000***
nodematch.Gender	-0.004988	0.100639	0	0.960476
nodeofactor.Gender.Male	0.22486	0.17008	0	0.186325
nodeifactor.Gender.Male	-0.649926	0.172313	0	0.000168***
nodematch.Education_Dich	0.127406	0.110063	0	0.24721
nodeocov.Education	-0.07652	0.076371	0	0.316513
nodeicov.Education	-0.076279	0.075568	0	0.312929
nodematch.Personality_Dich	-0.069842	0.093375	0	0.454581
nodeocov.Personality	0.319466	0.109572	0	0.003599**
nodeicov.Personality	-0.654279	0.108728	0	0.000000***
nodeocov.Prior_Contacts	0.042022	0.027264	0	0.123436
nodeicov.Prior_Contacts	-0.082171	0.028809	0	0.004396**

AIC: 1507 BIC: 1577

Table B.40. Bernoulli model results for EmailCo network (August 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.34117	0.06091	0	0.000000***

AIC: 1676 BIC: 1681

Table B.41. Model 1 results for EmailCo email network (August 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.79429	0.06287	0	0.000000***
gwidegree	-5.18048	0.30305	0	0.000000***

AIC: 1469 BIC: 1480

Table B.42. Model 2 results for EmailCo email network (August 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.922518	0.409921	0	0.000000***
mutual	3.961438	0.263883	0	0.000000***
nodematch.Gender	-0.041973	0.11577	0	0.716984
nodeofactor.Gender.Male	0.231956	0.194735	0	0.233776
nodeifactor.Gender.Male	-0.104782	0.192038	0	0.585393
nodematch.Education_Dich	-0.151308	0.112239	0	0.177817
nodeocov.Education	0.003422	0.087161	0	0.968691
nodeicov.Education	-0.155498	0.084392	0	0.065572.
nodematch.Personality_Dich	-0.014523	0.093576	0	0.876681
nodeocov.Personality	-0.328871	0.11932	0	0.005913**
nodeicov.Personality	0.30172	0.121965	0	0.013469*
nodeocov.Prior_Contacts	0.100925	0.026604	0	0.000154***
nodeicov.Prior_Contacts	-0.101005	0.031594	0	0.001416**

AIC: 1337 BIC: 1407

Table B.43. Bernoulli model results for EmailCo network (September 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.3081	0.06033	0	0.000000***

AIC: 1700 BIC: 1705

Table B.44. Model 1 results for EmailCo email network (September 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.74103	0.05786	0	0.000000***
gwidegree	-5.46608	0.30351	0	0.000000***

AIC: 1464 BIC: 1475

Table B.45. Model 2 results for EmailCo email network (September 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-3.176434	0.452279	0	0.000000***
mutual	4.063266	0.255682	0	0.000000***
nodematch.Gender	-0.042949	0.113465	0	0.705092
nodeofactor.Gender.Male	-0.111705	0.203046	0	0.582295
nodeifactor.Gender.Male	0.129606	0.199457	0	0.515918
nodematch.Education_Dich	-0.200797	0.125509	0	0.109825
nodeocov.Education	0.019744	0.09577	0	0.836695
nodeicov.Education	0.346326	0.097009	0	0.000367***
nodematch.Personality_Dich	-0.026753	0.098469	0	0.785896
nodeocov.Personality	-0.004893	0.124342	0	0.968616
nodeicov.Personality	-0.353702	0.122886	0	0.004050**
nodeocov.Prior_Contacts	0.029926	0.029594	0	0.312077
nodeicov.Prior_Contacts	-0.040645	0.031277	0	0.193954

AIC: 1320 BIC: 1390

Table B.46. Bernoulli model results for EmailCo network (October 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.36358	0.06132	0	0.000000***

AIC: 1660 BIC: 1665

Table B.47. Model 1 results for EmailCo email network (October 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.80579	0.05767	0	0.000000***
gwidegree	-5.14026	0.2997	0	0.000000***

AIC: 1450 BIC: 1461

Table B.48. Model 2 results for EmailCo email network (October 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.221688	0.446476	0	0.00628**
mutual	3.804102	0.256494	0	0.000000***
nodematch.Gender	-0.031841	0.10559	0	0.76303
nodeofactor.Gender.Male	-0.179115	0.183007	0	0.32786
nodeifactor.Gender.Male	-0.460712	0.182055	0	0.01148*
nodematch.Education_Dich	-0.009962	0.118642	0	0.93309
nodeocov.Education	-0.136674	0.086188	0	0.11299
nodeicov.Education	0.148478	0.087459	0	0.08976.
nodematch.Personality_Dich	-0.010274	0.101697	0	0.91954
nodeocov.Personality	0.217846	0.12291	0	0.07651.
nodeicov.Personality	-0.517026	0.117791	0	0.000000***
nodeocov.Prior_Contacts	0.013973	0.030628	0	0.6483
nodeicov.Prior_Contacts	-0.065288	0.031872	0	0.04068*

AIC: 1318 BIC: 1389

Table B.49. Bernoulli model results for EmailCo network (November 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.1314	0.0575	0	0.000000***

AIC: 1824 BIC: 1830

Table B.50. Model 1 results for EmailCo email network (November 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.7439	0.0575	0	0.000000***
gwidegree	-5.2377	0.3112	0	0.000000***

AIC: 1655 BIC: 1665

Table B.51. Model 2 results for EmailCo email network (November 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-3.33428	0.40897	0	0.000000***
mutual	3.70999	0.23557	0	0.000000***
nodematch.Gender	-0.03855	0.11448	0	0.736365
nodeofactor.Gender.Male	-0.26285	0.18236	0	0.149675
nodeifactor.Gender.Male	0.27273	0.18163	0	0.13339
nodematch.Education_Dich	0.15523	0.12135	0	0.201041
nodeocov.Education	0.1779	0.08665	0	0.040224*
nodeicov.Education	-0.02376	0.08812	0	0.787448
nodematch.Personality_Dich	-0.06473	0.09113	0	0.477644
nodeocov.Personality	0.05183	0.11432	0	0.650349
nodeicov.Personality	0.02062	0.11235	0	0.854403
nodeocov.Prior_Contacts	-0.01286	0.03092	0	0.677488
nodeicov.Prior_Contacts	-0.12257	0.03711	0	0.000978***

AIC: 1449 BIC: 1520

Table B.52. Bernoulli model results for EmailCo network (December 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.2936	0.06008	0	0.000000***

AIC: 1710 BIC: 1716

Table B.53. Model 1 results for EmailCo email network (December 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.86796	0.05786	0	0.000000***
gwidegree	-4.7272	0.28478	0	0.000000***

AIC: 1564 BIC: 1575

Table B.54. Model 2 results for EmailCo email network (December 2016)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-2.687968	0.418622	0	0.000000***
mutual	3.753148	0.238643	0	0.000000***
nodematch.Gender	-0.166517	0.114073	0	0.145
nodeofactor.Gender.Male	-0.113352	0.186752	0	0.544
nodeifactor.Gender.Male	0.243151	0.185035	0	0.189
nodematch.Education_Dich	0.04464	0.118489	0	0.706
nodeocov.Education	-0.00717	0.085889	0	0.933
nodeicov.Education	-0.021608	0.086546	0	0.803
nodematch.Personality_Dich	-0.008827	0.093003	0	0.924
nodeocov.Personality	0.001232	0.119741	0	0.992
nodeicov.Personality	0.078686	0.118515	0	0.507
nodeocov.Prior_Contacts	-0.009934	0.031479	0	0.752
nodeicov.Prior_Contacts	-0.044617	0.032982	0	0.176

AIC: 1388 BIC: 1458

Table B.55. Bernoulli model results for EmailCo network (January 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.46434	0.06323	0	0.000000***

AIC: 1586 BIC: 1592

Table B.56. Model 1 results for EmailCo email network (January 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.95493	0.06079	0	0.000000***
gwidegree	-4.42727	0.28475	0	0.000000***

AIC: 1437 BIC: 1448

Table B.57. Model 2 results for EmailCo email network (January 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-4.01359	0.44107	0	0.000000***
mutual	3.71227	0.25567	0	0.000000***
nodematch.Gender	0.11799	0.13614	0	0.386266
nodeofactor.Gender.Male	0.33013	0.20498	0	0.10747
nodeifactor.Gender.Male	0.32545	0.21162	0	0.124262
nodematch.Education_Dich	0.05971	0.12585	0	0.635229
nodeocov.Education	0.08611	0.09309	0	0.355094
nodeicov.Education	-0.01977	0.09366	0	0.832887
nodematch.Personality_Dich	-0.14229	0.10409	0	0.171816
nodeocov.Personality	-0.2728	0.1197	0	0.022788*
nodeicov.Personality	0.42999	0.12082	0	0.000383***
nodeocov.Prior_Contacts	0.0409	0.03045	0	0.179406
nodeicov.Prior_Contacts	-0.08989	0.03627	0	0.013293*

AIC: 1273 BIC: 1343

Table B.58. Bernoulli model results for EmailCo network (February 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.23327	0.05908	0	0.000000***

AIC: 1753 BIC: 1759

Table B.59. Model 1 results for EmailCo email network (February 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.8378	0.061	0	0.000000***
gwidegree	-4.8343	0.2983	0	0.000000***

AIC: 1610 BIC: 1621

Table B.60. Model 2 results for EmailCo email network (February 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-4.37061	0.44804	0	0.000000***
mutual	3.42166	0.2269	0	0.000000***
nodematch.Gender	-0.04391	0.11085	0	0.6921
nodeofactor.Gender.Male	0.21469	0.18033	0	0.234
nodeifactor.Gender.Male	-0.03035	0.1802	0	0.8663
nodematch.Education_Dich	0.12926	0.13371	0	0.3338
nodeocov.Education	0.19489	0.09243	0	0.0351*
nodeicov.Education	-0.01425	0.08752	0	0.8707
nodematch.Personality_Dich	-0.12131	0.09481	0	0.2009
nodeocov.Personality	0.16395	0.11007	0	0.1365
nodeicov.Personality	0.12042	0.11002	0	0.2739
nodeocov.Prior_Contacts	0.03966	0.02807	0	0.1579
nodeicov.Prior_Contacts	-0.07864	0.031	0	0.0113*

AIC: 1433 BIC: 1503

Table B.61. Bernoulli model results for EmailCo network (March 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.8299	0.0537	0	0.000000***

AIC: 2016 BIC: 2021

Table B.62. Model 1 results for EmailCo email network (March 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.52316	0.05459	0	0.000000***
gwidegree	-6.28263	0.33107	1	0.000000***

AIC: 1840 BIC: 1850

Table B.63. Model 2 results for EmailCo email network (March 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-2.262001	0.375615	0	0.000000***
mutual	3.278539	0.205394	0	0.000000***
nodematch.Gender	-0.134385	0.101834	0	0.1871
nodeofactor.Gender.Male	-0.206846	0.165112	0	0.2105
nodeifactor.Gender.Male	0.368316	0.160687	0	0.0220*
nodematch.Education_Dich	0.002903	0.110054	0	0.979
nodeocov.Education	0.081599	0.076378	0	0.2855
nodeicov.Education	-0.028742	0.075007	0	0.7016
nodematch.Personality_Dich	-0.076543	0.088049	0	0.3848
nodeocov.Personality	-0.032908	0.101077	0	0.7448
nodeicov.Personality	0.058673	0.099158	0	0.5541
nodeocov.Prior_Contacts	-0.056284	0.029714	0	0.0584.
nodeicov.Prior_Contacts	-0.070727	0.030643	0	0.0211*

AIC: 1662 BIC: 1732

Table B.64. Bernoulli model results for EmailCo network (April 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.98475	0.05549	0	0.000000***

AIC: 1921 BIC: 1927

Table B.65. Model 1 results for EmailCo email network (April 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.65012	0.05586	0	0.000000***
gwidegree	-5.67561	0.30935	0	0.000000***

AIC: 1759 BIC: 1770

Table B.66. Model 2 results for EmailCo email network (April 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-3.99841	0.386424	0	0.000000***
mutual	3.563458	0.223042	0	0.000000***
nodematch.Gender	0.106399	0.107128	0	0.320764
nodeofactor.Gender.Male	0.172383	0.180609	0	0.339995
nodeifactor.Gender.Male	0.059367	0.177338	0	0.737845
nodematch.Education_Dich	-0.014791	0.116626	0	0.899092
nodeocov.Education	0.116188	0.085298	0	0.173341
nodeicov.Education	0.078285	0.084826	0	0.356205
nodematch.Personality_Dich	-0.061987	0.087029	0	0.476409
nodeocov.Personality	0.007191	0.109329	0	0.947563
nodeicov.Personality	0.211142	0.109789	0	0.054635.
nodeocov.Prior_Contacts	-0.02258	0.032592	0	0.488529
nodeicov.Prior_Contacts	-0.117878	0.035565	0	0.000938***

AIC: 1516 BIC: 1587

Table B.67. Bernoulli model results for EmailCo network (May 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.34861	0.06105	0	0.000000***

AIC: 1671 BIC: 1676

Table B.68. Model 1 results for EmailCo email network (May 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.76417	0.05927	0	0.000000***
gwidegree	-5.35537	0.30183	0	0.000000***

AIC: 1437 BIC: 1447

Table B.69. Model 2 results for EmailCo email network (May 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-3.47007	0.42159	0	0.000000***
mutual	3.70521	0.24546	0	0.000000***
nodematch.Gender	0.01171	0.12078	0	0.9228
nodeofactor.Gender.Male	-0.08118	0.18884	0	0.6673
nodeifactor.Gender.Male	0.38649	0.19258	0	0.0449*
nodematch.Education_Dich	-0.14563	0.11971	0	0.224
nodeocov.Education	-0.08484	0.08489	0	0.3178
nodeicov.Education	0.17823	0.08697	0	0.0406*
nodematch.Personality_Dich	-0.06922	0.09701	0	0.4756
nodeocov.Personality	-0.01958	0.11594	0	0.8659
nodeicov.Personality	0.19146	0.11628	0	0.0998.
nodeocov.Prior_Contacts	-0.01587	0.03462	0	0.6468
nodeicov.Prior_Contacts	-0.18694	0.04332	0	0.000000***

AIC: 1324 BIC: 1394

Table B.70. Bernoulli model results for EmailCo network (June 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.56758	0.06535	0	0.000000***

AIC: 1511 BIC: 1516

Table B.71. Model 1 results for EmailCo email network (June 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.00066	0.06224	0	0.000000***
gwidegree	-4.30526	0.28251	0	0.000000***

AIC: 1359 BIC: 1370

Table B.72. Model 2 results for EmailCo email network (June 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-3.40654	0.44859	0	0.000000***
mutual	4.18139	0.2845	0	0.000000***
nodematch.Gender	-0.10169	0.114	0	0.372515
nodeofactor.Gender.Male	0.03166	0.20016	0	0.874356
nodeifactor.Gender.Male	-0.16367	0.20666	0	0.428477
nodematch.Education_Dich	0.28424	0.13033	0	0.029324*
nodeocov.Education	0.05773	0.10147	0	0.569457
nodeicov.Education	-0.03946	0.09925	0	0.691006
nodematch.Personality_Dich	0.08398	0.10968	0	0.443992
nodeocov.Personality	0.47204	0.13398	0	0.000438***
nodeicov.Personality	-0.41708	0.13387	0	0.001868**
nodeocov.Prior_Contacts	-0.04885	0.03227	0	0.130309
nodeicov.Prior_Contacts	0.07137	0.0293	0	0.014984*

AIC: 1193 BIC: 1264

Table B.73. Bernoulli model results for EmailCo network (July 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.38249	0.06166	0	0.000000***

AIC: 1646 BIC: 1651

Table B.74. Model 1 results for EmailCo email network (July 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.9159	0.05905	0	0.000000***
gwidegree	-4.5633	0.3013	0	0.000000***

AIC: 1499 BIC: 1510

Table B.75. Model 2 results for EmailCo email network (July 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-4.751906	0.522595	0	0.000000***
mutual	3.175849	0.230033	0	0.000000***
nodematch.Gender	-0.106249	0.136	0	0.43477
nodeofactor.Gender.Male	0.005934	0.191592	0	0.9753
nodeifactor.Gender.Male	0.556808	0.188799	0	0.00323**
nodematch.Education_Dich	-0.199797	0.146863	0	0.17388
nodeocov.Education	0.247856	0.094408	0	0.00874**
nodeicov.Education	0.17801	0.096174	0	0.06436.
nodematch.Personality_Dich	0.044737	0.100366	0	0.65584
nodeocov.Personality	-0.129233	0.110575	0	0.24268
nodeicov.Personality	0.097406	0.112047	0	0.38479
nodeocov.Prior_Contacts	-0.013362	0.027475	0	0.62681
nodeicov.Prior_Contacts	0.069307	0.026326	0	0.00855**

AIC: 1374 BIC: 1445

Table B.76. Bernoulli model results for EmailCo online network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.26141	0.05954	0	0.000000***

AIC: 1733 BIC: 1739

Table B.77. Model 1 results for EmailCo email network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.82089	0.05984	0	0.000000***
gwidegree	-4.93947	0.30254	0	0.000000***

AIC: 1569 BIC: 1579

Table B.78. Model 2 results for EmailCo email network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-2.95108	0.4093	0	0.000000***
mutual	3.80849	0.24192	0	0.000000***
nodematch.Gender	-0.0309	0.10559	0	0.7699
nodeofactor.Gender.Male	-0.44513	0.17771	0	0.0123*
nodeifactor.Gender.Male	-0.08862	0.17941	0	0.6214
nodematch.Education_Dich	0.12413	0.12066	0	0.3038
nodeocov.Education	0.03765	0.08888	0	0.6719
nodeicov.Education	-0.06366	0.08899	0	0.4745
nodematch.Personality_Dich	-0.01455	0.09597	0	0.8795
nodeocov.Personality	0.03023	0.11932	0	0.8
nodeicov.Personality	0.21597	0.11792	0	0.0672.
nodeocov.Prior_Contacts	-0.02713	0.03102	0	0.3818
nodeicov.Prior_Contacts	-0.03482	0.03193	0	0.2757

AIC: 1369 BIC: 1439

Table B.79. Bernoulli model results for EmailCo network (July 2015 - August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.8332	0.05653	0	0.000000***

AIC: 1820 BIC: 1825

Table B.80. Model 1 results for EmailCo email network (July 2015 - August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.71941	0.05666	0	0.000000***
gwidegree	-8.01755	0.6969	2	0.000000***

AIC: 1766 BIC: 1776

Table B.81. Model 2 results for EmailCo email network (July 2015 - August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.51017	0.39078	0	0.000116***
mutual	5.03652	0.31071	0	0.000000***
nodematch.Gender	-0.05053	0.09799	0	0.60615
nodeofactor.Gender.Male	-0.26	0.22982	0	0.258094
nodeifactor.Gender.Male	0.01673	0.22945	0	0.941891
nodematch.Education_Dich	0.05766	0.10469	0	0.581888
nodeocov.Education	0.01396	0.10356	0	0.89275
nodeicov.Education	-0.11622	0.10514	0	0.269174
nodematch.Personality_Dich	0.19719	0.09118	0	0.030735*
nodeocov.Personality	0.20942	0.14664	0	0.153466
nodeicov.Personality	-0.54231	0.14724	0	0.000239***
nodeocov.Prior_Contacts	0.03381	0.03617	0	0.350084
nodeicov.Prior_Contacts	0.06165	0.0358	0	0.085270.

AIC: 1243 BIC: 1312

Table B.82. Model 3 results for EmailCo email network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-3.573003	0.654159	0	0.000000***
mutual	3.929384	0.22932	0	0.000000***
nodematch.Gender	-0.007309	0.102207	0	0.942996
nodeofactor.Gender.Male	-0.667968	0.191337	0	0.000491***
nodeifactor.Gender.Male	-0.277588	0.194955	0	0.15464
nodematch.Education	0.101902	0.099859	0	0.307634
nodeocov.Education	0.014667	0.085517	0	0.863837
nodeicov.Education	0.073854	0.084532	0	0.382394
nodematch.Personality	0.140265	0.091778	0	0.12659
nodeocov.Personality	0.061157	0.122568	0	0.617862
nodeicov.Personality	0.556078	0.123137	0	0.000000***
nodeocov.Task_InterDep	0.147703	0.096402	0	0.125638
nodeicov.Task_InterDep	0.096056	0.096084	0	0.317568
nodeocov.Prior_Contacts	-0.019544	0.025601	0	0.445325
nodeicov.Prior_Contacts	-0.029599	0.02687	0	0.270789
nodeocov.Email	-0.106435	0.12917	0	0.410042
nodeicov.Email	-0.161836	0.127446	0	0.204286
nodeocov.Phone	-0.003627	0.083347	0	0.965295
nodeicov.Phone	-0.076684	0.08399	0	0.361346
nodeocov.F2F	-0.050653	0.084344	0	0.548201
nodeicov.F2F	-0.328406	0.083803	0	0.000000***
nodematch.Location	0.055921	0.099365	0	0.573647
nodefactor.Location.Not_HQ	0.388677	0.087215	0	0.000000***

AIC: 1599 BIC: 1729

Table B.83. Bernoulli model results for EmailCo advice network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.05041	0.05636	0	0.000000***

AIC: 1879 BIC: 1884

Table B.84. Model 1 results for EmailCo advice network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.82127	0.05579	0	0.000000***
gwidegree	-4.57004	0.31843	0	0.000000***

AIC: 1800 BIC: 1811

Table B.85. Model 3 results for EmailCo advice network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-5.342089	0.759662	0	0.000000***
mutual	1.914638	0.196143	0	0.000000***
nodematch.Gender	0.027204	0.11568	0	0.814108
nodeofactor.Gender.Male	0.588265	0.165238	0	0.000381***
nodeifactor.Gender.Male	-0.420257	0.158792	0	0.008210**
nodematch.Education	0.024938	0.108427	0	0.818126
nodeocov.Education	0.130146	0.072291	0	0.071997.
nodeicov.Education	0.058465	0.069545	0	0.400649
nodematch.Personality	0.157204	0.106177	0	0.138913
nodeocov.Personality	0.12187	0.101491	0	0.230005
nodeicov.Personality	-0.252182	0.102381	0	0.013874*
nodeocov.Task_InterDep	0.346929	0.07411	0	0.000000***
nodeicov.Task_InterDep	-0.01261	0.078073	0	0.871703
nodeocov.Prior_Contacts	-0.006951	0.023838	0	0.770651
nodeicov.Prior_Contacts	0.004083	0.022397	0	0.855374
nodeocov.Email	0.60633	0.109695	0	0.000000***
nodeicov.Email	-0.101997	0.105287	0	0.332813
nodeocov.Phone	-0.270372	0.067711	0	0.000000***
nodeicov.Phone	0.137238	0.066043	0	0.037867*
nodeocov.F2F	-0.246926	0.064179	0	0.000124***
nodeicov.F2F	0.144852	0.06734	0	0.031619*
nodematch.Location	-0.03011	0.103888	0	0.771984
nodefactor.Location.Not_HQ	-0.19531	0.0841	0	0.020337*

AIC: 1718 BIC: 184

Table B.86. Bernoulli model results for EmailCo brainstorming network (August 2017)

Terms	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.51717	0.06429	0	0.000000***

AIC: 1548 BIC: 1553

Table B.87. Model 1 results for EmailCo brainstorming network (August 2017)

Terms	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.10199	0.06275	0	0.000000***
gwidegree	-3.68898	0.2986	0	0.000000***

AIC: 1463 BIC: 1474

Table B.88. Model 3 results for EmailCo brainstorming network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-5.88299	0.83538	0	0.000000***
mutual	1.96133	0.21904	0	0.000000***
nodematch.Gender	0.07678	0.13338	0	0.564952
nodeofactor.Gender.Male	0.6693	0.19062	0	0.000458***
nodeifactor.Gender.Male	-0.35797	0.18101	0	0.048139*
nodematch.Education	-0.13475	0.12761	0	0.291165
nodeocov.Education	0.04692	0.07906	0	0.552927
nodeicov.Education	0.08893	0.07638	0	0.244435
nodematch.Personality	0.05872	0.12322	0	0.633745
nodeocov.Personality	-0.06535	0.11054	0	0.55445
nodeicov.Personality	-0.1937	0.112	0	0.083911.
nodeocov.Task_InterDep	0.2017	0.08284	0	0.015010*
nodeicov.Task_InterDep	0.06178	0.0828	0	0.455694
nodeocov.Prior_Contacts	-0.02051	0.03046	0	0.500733
nodeicov.Prior_Contacts	-0.08136	0.03165	0	0.010243*
nodeocov.Email	0.54348	0.11994	0	0.000000***
nodeicov.Email	0.15117	0.11843	0	0.201994
nodeocov.Phone	-0.26439	0.07671	0	0.000582***
nodeicov.Phone	0.11572	0.07336	0	0.114882
nodeocov.F2F	0.05088	0.07318	0	0.486964
nodeicov.F2F	-0.01203	0.07224	0	0.867774
nodematch.Location	-0.06609	0.1169	0	0.571906
nodefactor.Location.Not_HQ	-0.09404	0.09552	0	0.325024

AIC: 1433 BIC: 1557

Table B.89. Bernoulli model results for EmailCo working network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-1.138	0.0576	0	0.000000***

AIC: 1820 BIC: 1825

Table B.90. Model 1 results for EmailCo working network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-0.87976	0.05635	0	0.000000***
gwidegree	-4.38414	0.30015	0	0.000000***

AIC: 1741 BIC: 1751

Table B.91. Model 3 results for EmailCo working network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-4.961676	0.740226	0	0.000000***
mutual	2.790445	0.223415	0	0.000000***
nodematch.Gender	0.108446	0.110466	0	0.326386
nodeofactor.Gender.Male	0.902243	0.183715	0	0.000000***
nodeifactor.Gender.Male	-0.737732	0.17721	0	0.000000***
nodematch.Education	-0.019398	0.111496	0	0.861901
nodeocov.Education	-0.078693	0.076148	0	0.301564
nodeicov.Education	0.21108	0.075353	0	0.005152**
nodematch.Personality	0.132109	0.105259	0	0.20963
nodeocov.Personality	-0.024947	0.106683	0	0.815133
nodeicov.Personality	-0.357942	0.104477	0	0.000628***
nodeocov.Task_InterDep	0.304474	0.079067	0	0.000122***
nodeicov.Task_InterDep	-0.112316	0.08452	0	0.184079
nodeocov.Prior_Contacts	-0.030931	0.027407	0	0.259239
nodeicov.Prior_Contacts	0.001992	0.025375	0	0.937425
nodeocov.Email	0.844597	0.122898	0	0.000000***
nodeicov.Email	-0.171658	0.116981	0	0.14246
nodeocov.Phone	-0.25567	0.075424	0	0.000716***
nodeicov.Phone	0.135104	0.071757	0	0.059906.
nodeocov.F2F	-0.089368	0.068762	0	0.1939
nodeicov.F2F	0.078334	0.071899	0	0.276096
nodematch.Location	-0.137541	0.100103	0	0.169635
nodefactor.Location.Not_HQ	-0.142118	0.084208	0	0.091660.

AIC: 1560 BIC: 168

Table B.92. Model 4 results for EmailCo email network (August 2017)

Term	*Estimate*	*StdError*	*MCMC%*	*p-value*
edges	-3.018644	0.716316	0	0.000000***
mutual	3.976964	0.234386	0	0.000000***
nodematch.Gender	-0.012705	0.097214	0	0.89603
nodeofactor.Gender.Male	-0.845685	0.207588	0	0.000000***
nodeifactor.Gender.Male	-0.358276	0.209388	0	0.087222.
nodematch.Education	0.046704	0.096608	0	0.628838
nodeocov.Education	0.005198	0.089469	0	0.953676
nodeicov.Education	0.063757	0.087547	0	0.466539
nodematch.Personality	0.116661	0.098639	0	0.237063
nodeocov.Personality	0.072752	0.126296	0	0.564648
nodeicov.Personality	0.474615	0.128378	0	0.000224***
nodeocov.Task_InterDep	0.149005	0.09617	0	0.121442
nodeicov.Task_InterDep	0.139202	0.095398	0	0.144671
nodeocov.Prior_Contacts	-0.010009	0.025647	0	0.696395
nodeicov.Prior_Contacts	-0.045559	0.027264	0	0.094864.
nodeocov.Email	-0.072822	0.131939	0	0.581054
nodeicov.Email	-0.162913	0.129105	0	0.207143
nodeocov.Phone	-0.059199	0.08461	0	0.484211
nodeicov.Phone	-0.132295	0.08412	0	0.115944
nodeocov.F2F	-0.037331	0.091899	0	0.684624
nodeicov.F2F	-0.336965	0.0873	0	0.000117***
nodematch.Location	0.061317	0.102699	0	0.550534
nodefactor.Location.Not_HQ	0.366434	0.087235	0	0.000000***
nodeicov.AdvIDegree	0.017895	0.011343	0	0.114801
nodeicov.InnoIDegree	0.022557	0.013915	0	0.105149
nodeicov.WorkIDegree	-0.040776	0.010939	0	0.000199***
nodeocov.AdvODegree	-0.02975	0.008913	0	0.000860***
nodeocov.InnoODegree	0.011819	0.010597	0	0.264843
nodeocov.WorkODegree	0.026978	0.010269	0	0.008679**

AIC: 1579 BIC: 1702

BIBLIOGRAPHY

Agarwal, Ritu, Anil K. Gupta, and Robert Kraut. 2008. "Editorial Overview-The Interplay between Digital and Social Networks." *Information Systems Research* 19 (3): 243–252.

Agneessens, Filip, and Rafael Wittek. 2008. "Social Capital and Employee Well-Being: Disentangling Intrapersonal and Interpersonal Selection and Influence Mechanisms." *Revue Française de Sociologie* 49 (3): 613–637.

Agrawal, Divyakant, Bassam Bamieh, Ceren Budak, Amr El Abbadi, Andrew Flanagin, and Stacy Patterson. 2011. "Data-Driven Modeling and Analysis of Online Social Networks." In *International Conference on Web-Age Information Management*, 3–17.

Ahuja, Gautam. 2000. "Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study." *Administrative Science Quarterly* 45 (3): 425–455.

Akaike, Hirotugu. 1974. "A New Look at the Statistical Model Identification." *Automatic Control, IEEE Transactions On* 19 (6): 716–723.

Albert, Réka, Hawoong Jeong, and Albert-László Barabási. 1999. "The Diameter of the World Wide Web." *ArXiv Preprint Cond-Mat/9907038*.

Bapna, Ravi, Alok Gupta, Sarah Rice, and Arun Sundararajan. 2011. "Trust, Reciprocity and the Strength of Social Ties: An Online Social Network Based Field Experiment." In *Conference on Information Systems and Technology (CIST)*.

Barabási, Albert-László, and Eric Bonabeau. 2003. "Scale-Free." *Scientific American*. http://eaton.math.rpi.edu/csums/papers/FoodWebs/barabasisciam.pdf.

Bavelas, Alex. 1948. "A Mathematical Model for Group Structures." *Human Organization* 7 (3): 16–30.

Beek, Adriana PA van, Cordula Wagner, Peter PM Spreeuwenberg, Dinnus HM Frijters, Miel W. Ribbe, and Peter P. Groenewegen. 2011. "Communication, Advice Exchange and Job Satisfaction of Nursing Staff: A Social Network Analyses of 35 Long-Term Care Units." *BMC Health Services Research* 11 (1): 140.

Benbasat, Izak, and Weiquan Wang. 2005. "Trust in and Adoption of Online Recommendation Agents." *Journal of the Association for Information Systems* 6 (3): 4.

Bickart, Kevin C., Christopher I. Wright, Rebecca J. Dautoff, Bradford C. Dickerson, and Lisa Feldman Barrett. 2011. "Amygdala Volume and Social Network Size in Humans." *Nature Neuroscience* 14 (2): 163–164.

Blau, Peter M. 1977. "A Macrosociological Theory of Social Structure." *American Journal of Sociology* 83 (1): 26–54.

Boivie, Steven, Michael K. Bednar, and Steven B. Barker. 2015. "Social Comparison and Reciprocity in Director Compensation." *Journal of Management* 41 (6): 1578–1603. https://doi.org/10.1177/0149206312460680.

Bonacich, Phillip. 1972. "Factoring and Weighting Approaches to Status Scores and Clique Identification." *Journal of Mathematical Sociology* 2 (1): 113–120.

———. 2007. "Some Unique Properties of Eigenvector Centrality." *Social Networks* 29 (4): 555–564.

Bordia, Prashant, Elizabeth Jones, Cindy Gallois, Victor J. Callan, and Nicholas DiFonzo. 2006. "Management Are Aliens! Rumors and Stress during Organizational Change." *Group & Organization Management* 31 (5): 601–621.

Borgatti, Stephen P. 1995. "Centrality and AIDS." *Connections* 18 (1): 112–114.

Borgatti, Stephen P., Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. 2009. "Network Analysis in the Social Sciences." *Science* 323 (5916): 892–895.

Bourdieu, Pierre. 2011. "The Forms of Capital.(1986)." *Cultural Theory: An Anthology* 1: 81–93.

Brass, Daniel J., Kenneth D. Butterfield, and Bruce C. Skaggs. 1998. "Relationships and Unethical Behavior: A Social Network Perspective." *Academy of Management Review* 23 (1): 14–31.

Brown, Deborah Wright, and Alison M. Konrad. 2001. "Granovetter Was Right: The Importance of Weak Ties to a Contemporary Job Search." *Group & Organization Management* 26 (4): 434–462.

Burr, Ronald S. 1992. "Structural Holes." *The Social Structure of Competition*.

Burt, Ronald S. 2002. "The Social Capital of Structural Holes." *The New Economic Sociology: Developments in an Emerging Field*, 148–190.

———. 2004. "Structural Holes and Good Ideas." *American Journal of Sociology* 110 (2): 349–399.

Buunk, Bram P., Esther S. Kluwer, Mieke K. Schuurman, and Frans W. Siero. 2000. "The Division of Labor among Egalitarian and Traditional Women: Differences in Discontent, Social Comparison, and False Consensus." *Journal of Applied Social Psychology* 30 (4): 759–779.

Carmi, Eyal, Gal Oestreicher-Singer, and Arun Sundararajan. 2012. "Is Oprah Contagious? Identifying Demand Spillovers in Online Networks." *Identifying Demand Spillovers in Online Networks (August 3, 2012) .NET Institute Working Paper*, no. 10–18. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1694308.

Carnabuci, Gianluca, and Bálint Diószegi. 2015. "Social Networks, Cognitive Style, and Innovative Performance: A Contingency Perspective." *Academy of Management Journal* 58 (3): 881–905.

Chiu, Chao-Min, Meng-Hsiang Hsu, and Eric T. G. Wang. 2006. "Understanding Knowledge Sharing in Virtual Communities: An Integration of Social Capital and Social Cognitive Theories." *Decision Support Systems* 42 (3): 1872–88. https://doi.org/10.1016/j.dss.2006.04.001.

Chow, Wing S., and Lai Sheung Chan. 2008. "Social Network, Social Trust and Shared Goals in Organizational Knowledge Sharing." *Information & Management* 45 (7): 458–465.

Christakis, Nicholas A., and James H. Fowler. 2007. "The Spread of Obesity in a Large Social Network over 32 Years." *New England Journal of Medicine* 357 (4): 370–79. https://doi.org/10.1056/NEJMsa066082.

———. 2008. "The Collective Dynamics of Smoking in a Large Social Network." *New England Journal of Medicine* 358 (21): 2249–2258.

———. 2013. "Social Contagion Theory: Examining Dynamic Social Networks and Human Behavior." *Statistics in Medicine* 32 (4): 556–77. https://doi.org/10.1002/sim.5408.

Cohen, Allan, and David Bradford. 2007. *Influence Without Authority 2Nd Ed.* John Wiley & Sons.

Cohn, Bernard S., and McKim Marriott. 1958. "Networks and Centres of Integration in Indian Civilization." *Journal of Social Research* 1 (1): 1–9.

Coleman, James S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology*, S95–S120.

Coles, Nigel. 2001. "It's Not What You Know—It's Who You Know That Counts. Analysing Serious Crime Groups as Social Networks." *British Journal of Criminology* 41 (4): 580–594.

Contractor, Noshir S., and Eric M. Eisenberg. 1990. "Communication Networks and New Media in Organizations." *Organizations and Communication Technology* 143: 172.

Cross, Rob, Stephen P. Borgatti, and Andrew Parker. 2001. "Beyond Answers: Dimensions of the Advice Network." *Social Networks* 23 (3): 215–235.

Csardi, Gabor, and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." *InterJournal, Complex Systems* 1695 (5): 1–9.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin. 2009. "An Economic Model of Friendship: Homophily, Minorities, and Segregation." *Econometrica* 77 (4): 1003–1045.

Diesner, Jana, Terrill L. Frantz, and Kathleen M. Carley. 2005. "Communication Networks from the Enron Email Corpus 'It's Always about the People. Enron Is No Different.'" *Computational & Mathematical Organization Theory* 11 (3): 201–228.

DiFonzo, Nicholas, and Prashant Bordia. 2007. *Rumor Psychology: Social and Organizational Approaches.* American Psychological Association.

DiFonzo, Nicholas, Prashant Bordia, and Ralph L. Rosnow. 1994. "Reining in Rumors." *Organizational Dynamics* 23 (1): 47–62.

DiFonzo, Nicholas, Martin J. Bourgeois, Jerry Suls, Christopher Homan, Noah Stupak, Bernard P. Brooks, David S. Ross, and Prashant Bordia. 2013. "Rumor Clustering, Consensus, and Polarization: Dynamic Social Impact and Self-Organization of Hearsay." *Journal of Experimental Social Psychology* 49 (3): 378–99. https://doi.org/10.1016/j.jesp.2012.12.010.

Dissanayake, Indika, Jie Zhang, and Bin Gu. 2015. "Task Division for Team Success in Crowdsourcing Contests: Resource Allocation and Alignment Effects." *Journal of Management Information Systems* 32 (2): 8–39.

Dou, Yifan, Marius F. Niculescu, and D. J. Wu. 2013. "Engineering Optimal Network Effects via Social Media Features and Seeding in Markets for Digital Goods and Services." *Information Systems Research* 24 (1): 164–185.

Dunbar, Robin IM. 1992. "Neocortex Size as a Constraint on Group Size in Primates." *Journal of Human Evolution* 22 (6): 469–493.

Egorov, Georgy and Mattias Polborn. 2010. "An Informational Theory of Homophily." http://www.econ.upf.edu/docs/seminars/egorov.pdf.

Erdős, P., and A. Renyi. 1959. "On Random Graphs I." *Publ. Math. Debrecen* 6: 290–297.

Eysenck, Hans Jurgen, and S. G. B. Eysenck. 1965. "The Eysenck Personality Inventory."

Fang, Xiao, Paul Jen-Hwa Hu, Zhepeng Li, and Weiyu Tsai. 2013. "Predicting Adoption Probabilities in Social Networks." *Information Systems Research* 24 (1): 128–145.

Faraj, Samer, and Steven L. Johnson. 2011. "Network Exchange Patterns in Online Communities." *Organization Science* 22 (6): 1464–1480.

Faraj, Samer, Srinivas Kudaravalli, and Molly Wasko. 2015. "Leading Collaboration in Online Communities." *Mis Quarterly* 39 (2): 393–412.

Feiler, Daniel C., and Adam M. Kleinbaum. 2015. "Popularity, Similarity, and the Network Extraversion Bias." *Psychological Science* 26 (5): 593–603.

Feller, Joseph, Patrick Finnegan, Brian Fitzgerald, and Jeremy Hayes. 2008. "From Peer Production to Productization: A Study of Socially Enabled Business Exchanges in Open Source Service Networks." *Information Systems Research* 19 (4): 475–493.

Fernandez, Roberto M., and Roger V. Gould. 1994. "A Dilemma of State Power: Brokerage and Influence in the National Health Policy Domain." *American Journal of Sociology* 99 (6): 1455–1491.

Festinger, Leon. 1954. "A Theory of Social Comparison Processes." *Human Relations* 7 (2): 117–140.

Fine, Gary Alan, and Patricia A. Turner. 2001. *Whispers on the Color Line: Rumor and Race in America*. Univ of California Press.

Fischer, Ronald. 2004. "Organizational Reward Allocation: A Comparison of British and German Organizations." *International Journal of Intercultural Relations* 28 (2): 151–164.

Fowler, James H., and Nicholas A. Christakis. 2008. "Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis over 20 Years in the Framingham Heart Study." *Bmj* 337: a2338.

Fowler, James H., Christopher T. Dawes, and Nicholas A. Christakis. 2009. "Model of Genetic Variation in Human Social Networks." *Proceedings of the National Academy of Sciences* 106 (6): 1720–1724.

Freeman, Linton C. 1977. “A Set of Measures of Centrality Based on Betweenness.” *Sociometry*, 35–41.

———. 1978. “Centrality in Social Networks Conceptual Clarification.” *Social Networks* 1 (3): 215–239.

Freeman, Richard B., and Wei Huang. 2015. “Collaborating with People like Me: Ethnic Coauthorship within the United States.” *Journal of Labor Economics* 33 (S1): S289–S318.

Gächter, Simon, Daniele Nosenzo, and Martin Sefton. 2012. “The Impact of Social Comparisons on Reciprocity.” *The Scandinavian Journal of Economics* 114 (4): 1346–1367.

Gallivan, Michael, and Manju Ahuja. 2015. “Co-Authorship, Homophily, and Scholarly Influence in Information Systems Research.” *Journal of the Association for Information Systems* 16 (12): 980.

Gattis, Krista S., Sara Berns, Lorelei E. Simpson, and Andrew Christensen. 2004. “Birds of a Feather or Strange Birds? Ties Among Personality Dimensions, Similarity, and Marital Quality.” *Journal of Family Psychology* 18 (4): 564–74. https://doi.org/10.1037/0893-3200.18.4.564.

Gibbons, Deborah E. 2004. “Friendship and Advice Networks in the Context of Changing Professional Values.” *Administrative Science Quarterly* 49 (2): 238–262.

Gloor, Peter A. 2016. “What Email Reveals About Your Organization.” *MIT Sloan Management Review* 57 (2): 8.

Gnyawali, Devi R., Weiguo Fan, and James Penner. 2010. “Competitive Actions and Dynamics in the Digital Age: An Empirical Investigation of Social Networking Firms.” *Information Systems Research* 21 (3): 594–613.

Godde, Sophie, Lionel Humbert, Steeve D. Côté, Denis Réale, and Hal Whitehead. 2013. “Correcting for the Impact of Gregariousness in Social Network Analyses.” *Animal Behaviour* 85 (3): 553–558.

Godinho de Matos, Miguel, Pedro Ferreira, and David Krackhardt. 2014. “Peer Influence in the Diffusion of the IPhone 3G over a Large Social Network.” *Management Information Systems Quarterly (Forthcoming)*.

Goodreau, Steven M. 2007. “Advances in Exponential Random Graph (P*) Models Applied to a Large Social Network.” *Social Networks* 29 (2): 231–48. https://doi.org/10.1016/j.socnet.2006.08.001.

Granovetter, Mark. 1983. “The Strength of Weak Ties: A Network Theory Revisited.” *Sociological Theory* 1 (1): 201–233.

Granovetter, Mark S. 1973. “The Strength of Weak Ties.” *American Journal of Sociology*, 1360–1380.

Gray, Peter H., Salvatore Parise, and Bala Iyer. 2011. “Innovation Impacts of Using Social Bookmarking Systems.” *MIS Quarterly* 35 (3): 629–643.

Guo, Hong, H. Kenneth Cheng, and Ken Kelley. 2015. “Impact of Network Structure on Malware Propagation: A Growth Curve Perspective.” *Available at SSRN 1286311*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1286311.

Guzzo, Richard A., and Gregory P. Shea. 1992. “Group Performance and Intergroup Relations in Organizations.”

Hahn, Jungpil, Jae Yun Moon, and Chen Zhang. 2008. “Emergence of New Project Teams from Open Source Software Developer Networks: Impact of Prior Collaboration Ties.” *Information Systems Research* 19 (3): 369–391.

Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2008. “Statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data.” *Journal of Statistical Software* 24 (1): 1548.

Hanneman, Robert A., and Mark Riddle. 2005. *Introduction to Social Network Methods*. University of California Riverside.

Hansen, Morten T. 1999. “The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits.” *Administrative Science Quarterly* 44 (1): 82–111.

Harris, Jenine K. 2013. *An Introduction to Exponential Random Graph Modeling*. Vol. 173. Sage Publications.

Hayes, Natalie, and Stephen Joseph. 2003. “Big 5 Correlates of Three Measures of Subjective Well-Being.” *Personality and Individual Differences* 34 (4): 723–27. https://doi.org/10.1016/S0191-8869(02)00057-0.

Herring, Susan C. 2008. “9 Gender and Power in On–line Communication.” *The Handbook of Language and Gender* 25: 202.

Hinz, Oliver, and Martin Spann. 2008. “The Impact of Information Diffusion on Bidding Behavior in Secret Reserve Price Auctions.” *Information Systems Research* 19 (3): 351–368.

Howison, James, Andrea Wiggins, and Kevin Crowston. 2011. "Validity Issues in the Use of Social Network Analysis with Digital Trace Data." *Journal of the Association for Information Systems* 12 (12): 767.

Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2008. "Ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks." *Journal of Statistical Software* 24 (3): nihpa54860.

Ibarra, Herminia. 1992. "Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm." *Administrative Science Quarterly*, 422–447.

———. 1995. "Race, Opportunity, and Diversity of Social Circles in Managerial Networks." *Academy of Management Journal* 38 (3): 673–703.

Ihaka, Ross, and Robert Gentleman. 1996. "R: A Language for Data Analysis and Graphics." *Journal of Computational and Graphical Statistics* 5 (3): 299–314.

Jackman, Simon, and Lynn Vavreck. 2010. "Primary Politics: Race, Gender, and Age in the 2008 Democratic Primary." *Journal of Elections, Public Opinion and Parties* 20 (2): 153–186.

Johnson, Steven L., Samer Faraj, and Srinivas Kudaravalli. 2014. "Emergence of Power Laws in Online Communities: The Role of Social Mechanisms and Preferential Attachment." *Mis Quarterly* 38 (3): 795–808.

Kalmijn, Matthijs. 1998. "Intermarriage and Homogamy: Causes, Patterns, Trends." *Annual Review of Sociology* 24 (1): 395–421.

Kane, Gerald C., and Maryam Alavi. 2008. "Casting the Net: A Multimodal Network Perspective on User-System Interactions." *Information Systems Research* 19 (3): 253–272.

Kane, Gerald C., Maryam Alavi, Giuseppe Joe Labianca, and Steve Borgatti. 2012. "What's Different about Social Media Networks? A Framework and Research Agenda." *MIS Quarterly, Forthcoming*.

Kane, Gerald C., and Steve Borgatti. 2011. "Centrality–IS Proficiency Alignment and Workgroup Performance." https://works.bepress.com/steveborgatti/1/.

Kane, Gerald C., and Sam Ransbotham. 2016. "Research Note—Content and Collaboration: An Affiliation Network Approach to Information Quality in Online Peer Production Communities." *Information Systems Research*. http://pubsonline.informs.org/doi/abs/10.1287/isre.2016.0622.

Kardos, Peter, Bernhard Leidner, Csaba Pléh, Péter Soltész, and Zsolt Unoka. 2017. "Empathic People Have More Friends: Empathic Abilities Predict Social Network Size and Position in Social Network Predicts Empathic Efforts." *Social Networks* 50: 1–5.

Kilduff, Martin, and Wenpin Tsai. 2003. *Social Networks and Organizations*. Sage.

Klein, Katherine J., Beng-Chong Lim, Jessica L. Saltz, and David M. Mayer. 2004. "How Do They Get There? An Examination of the Antecedents of Centrality in Team Networks." *Academy of Management Journal* 47 (6): 952–963.

Koschützki, Dirk, Katharina Anna Lehmann, Leon Peeters, Stefan Richter, Dagmar Tenfelde-Podehl, and Oliver Zlotowski. 2005. "Centrality Indices." In *Network Analysis*, 16–61. Springer.

Kramer, Adam DI, Jamie E. Guillory, and Jeffrey T. Hancock. 2014. "Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks." *Proceedings of the National Academy of Sciences* 111 (24): 8788–8790.

Kudaravalli, Srinivas, and Samer Faraj. 2008. "The Structure of Collaboration in Electronic Networks." *Journal of the Association for Information Systems* 9 (10/11): 706.

Kudo, H., and R. I. M. Dunbar. 2001. "Neocortex Size and Social Network Size in Primates." *Animal Behaviour* 62 (4): 711–722.

Langfred, Claus W. 2005. "Autonomy and Performance in Teams: The Multilevel Moderating Effect of Task Interdependence." *Journal of Management* 31 (4): 513–29. https://doi.org/10.1177/0149206304272190.

Lazarsfeld, Paul F., Robert K. Merton, and others. 1954. "Friendship as a Social Process: A Substantive and Methodological Analysis." *Freedom and Control in Modern Society* 18 (1): 18–66.

Leonardi, Paul M. 2013. "When Does Technology Use Enable Network Change in Organizations? A Comparative Study of Feature Use and Shared Affordances." *Mis Quarterly* 37 (3): 749–775.

———. 2015. "Ambient Awareness and Knowledge Acquisition: Using Social Media to Learn 'Who Knows What' and 'Who Knows Whom.'" *Mis Quarterly* 39 (4): 747–762.

Levina, Natalia, and Manuel Arriaga. 2014. "Distinction and Status Production on User-Generated Content Platforms: Using Bourdieu's Theory of Cultural Production to Understand Social Dynamics in Online Fields." *Information Systems Research* 25 (3): 468–488.

Lin, Nan. 1999. “Building a Network Theory of Social Capital.” *Connections* 22 (1): 28–51.

Malinowski, Bronislaw. 1954. “Magic, Science and Religion.” *New York*, 85–87.

———. 2002. *Argonauts of the Western Pacific: An Account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Guinea*. Routledge.

Marsden, Peter V. 1988. “Homogeneity in Confiding Relations.” *Social Networks* 10 (1): 57–76.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology*, 415–444.

Merten, Frank, and Peter Gloor. 2010. “Too Much E-Mail Decreases Job Satisfaction.” *Procedia-Social and Behavioral Sciences* 2 (4): 6457–6465.

Moreno, Jacob L. 1934. *Who Shall Survive*. Vol. 58. JSTOR.

Moreno, Megan A., Lauren A. Jelenchick, Katie G. Egan, Elizabeth Cox, Henry Young, Kerry E. Gannon, and Tara Becker. 2011. “Feeling Bad on Facebook: Depression Disclosures by College Students on a Social Networking Site.” *Depression and Anxiety* 28 (6): 447–455.

Morris, Martina, Mark S. Handcock, and David R. Hunter. 2008. “Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects.” *Journal of Statistical Software* 24 (4): 1548–7660.

Muchnik, Lev, Sen Pei, Lucas C. Parra, Saulo D. S. Reis, José S. Andrade Jr, Shlomo Havlin, and Hernán A. Makse. 2013. “Origins of Power-Law Degree Distribution in the Heterogeneity of Human Activity in Social Networks.” *Scientific Reports* 3 (May): 1783. https://doi.org/10.1038/srep01783.

Newman, Mark. 2001a. “The Structure of Scientific Collaboration Networks.” *Proceedings of the National Academy of Sciences* 98 (2): 404–9. https://doi.org/10.1073/pnas.98.2.404.

———. 2001b. “Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality.” *Physical Review E* 64 (1). https://doi.org/10.1103/PhysRevE.64.016132.

Nieminen, Juhani. 1974. “On the Centrality in a Graph.” *Scandinavian Journal of Psychology* 15 (1): 332–336.

Oinas-Kukkonen, Harri, Kalle Lyytinen, and Youngjin Yoo. 2010. "Social Networks and Information Systems: Ongoing and Future Research Streams." *Journal of the Association for Information Systems* 11 (2): 3.

Okamoto, Kazuya, Wei Chen, and Xiang-Yang Li. 2008. "Ranking of Closeness Centrality for Large-Scale Social Networks." *Lecture Notes in Computer Science* 5059: 186–195.

Pan, Shan L., Gary Pan, and Dorothy E. Leidner. 2012. "Crisis Response Information Networks." *Journal of the Association for Information Systems* 13 (1): 31.

Patacchini, Eleonora, and Yves Zenou. 2008. "The Strength of Weak Ties in Crime." *European Economic Review* 52 (2): 209–236.

Phillips, James G., and Linnea Reddie. 2007. "Decisional Style and Self-Reported Email Use in the Workplace." *Computers in Human Behavior* 23 (5): 2414–28. https://doi.org/10.1016/j.chb.2006.03.016.

Polites, Greta L., and Richard T. Watson. 2009. "Using Social Network Analysis to Analyze Relationships among IS Journals." *Journal of the Association for Information Systems* 10 (8): 2.

Putzke, Johannes, Kai Fischbach, Detlef Schoder, and Peter Gloor. 2010. "The Evolution of Interaction Networks in Massively Multiplayer Online Games." *Journal of the Association for Information Systems* 11 (2).

Qiu, Liangfei, Huaxia Rui, and Andrew B. Whinston. 2014. "The Impact of Social Network Structures on Prediction Market Accuracy in the Presence of Insider Information." *Journal of Management Information Systems* 31 (1): 145–172.

Ransbotham, Sam, and Gerald C. Kane. 2011. "Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia." *MIS Quarterly-Management Information Systems* 35 (3): 613.

Reagans, Ray, and Ezra W. Zuckerman. 2001. "Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams." *Organization Science* 12 (4): 502–517.

Reynolds, Katherine J. 1987. "Self-Categorization Theory." *The Wiley-Blackwell Encyclopedia of Social Theory*.

Ridings, Catherine, and Molly Wasko. 2010. "Online Discussion Group Sustainability: Investigating the Interplay between Structural Dynamics and Social Dynamics over Time." *Journal of the Association for Information Systems* 11 (2): 95.

Robert Jr, Lionel P., Alan R. Dennis, and Manju K. Ahuja. 2008. "Social Capital and Knowledge Integration in Digitally Enabled Teams." *Information Systems Research* 19 (3): 314–334.

Robins, Garry, and Martina Morris. 2007. "Advances in Exponential Random Graph (P*) Models." *Social Networks* 29 (2): 169–72.

Santos, J. Reynaldo A. 1999. "Cronbach's Alpha: A Tool for Assessing the Reliability of Scales." *Journal of Extension* 37 (2): 1–5.

Sato, Toru. 2005. "The Eysenck Personality Questionnaire Brief Version: Factor Structure and Reliability." *The Journal of Psychology* 139 (6): 545–52. https://doi.org/10.3200/JRLP.139.6.545-552.

Shi, Zhan, Huaxia Rui, and Andrew B. Whinston. 2013. "Content Sharing in a Social Broadcasting Environment: Evidence from Twitter." *Available at SSRN 2341243*. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2341243.

Simmel, Georg. 1903. *The Metropolis and Mental Life*. na.

Suri, Siddharth, and Duncan J. Watts. 2011. "Cooperation and Contagion in Web-Based, Networked Public Goods Experiments." *PloS One* 6 (3): e16836.

Susarla, Anjana, Jeong-Ha Oh, and Yong Tan. 2012. "Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube." *Information Systems Research* 23 (1): 23–41.

Sykes, Tracy Ann, Viswanath Venkatesh, and Sanjay Gosain. 2009. "Model of Acceptance with Peer Support: A Social Network Perspective to Understand Employees' System Use." *MIS Quarterly*, 371–393.

Sykes, Tracy Ann, Viswanath Venkatesh, and Jonathan L. Johnson. 2014. "Enterprise System Implementation and Employee Job Performance: Understanding the Role of Advice Networks." *Mis Quarterly* 38 (1): 51–72.

Thaden, Lyssa L., and Thomas Rotolo. 2009. "The Measurement of Social Networks: A Comparison of Alter-Centered and Relationship-Centered Survey Designs." *Connections* 29 (1): 15–25.

Tichy, Noel M., Michael L. Tushman, and Charles Fombrun. 1979. "Social Network Analysis for Organizations." *Academy of Management Review* 4 (4): 507–519.

Travers, Jeffrey, and Stanley Milgram. 1969. "An Experimental Study of the Small World Problem." *Sociometry*, 425–443.

Tsui, Anne S., Jone L. Pearce, Lyman W. Porter, and Angela M. Tripoli. 1997. "Alternative Approaches to the Employee-Organization Relationship: Does Investment in Employees Pay Off?" *Academy of Management Journal* 40 (5): 1089–1121.

Van Tilburg, Theo. 1998. "Losing and Gaining in Old Age: Changes in Personal Network Size and Social Support in a Four-Year Longitudinal Study." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 53 (6): S313–S323.

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Vol. 8. Cambridge university press.

Watts, Alison. 2003. "A Dynamic Model of Network Formation." In *Networks and Groups*, 337–345. Springer. http://link.springer.com/chapter/10.1007/978-3-540-24790-6_15.

Wu, Lynn. 2013. "Social Network Effects on Productivity and Job Security: Evidence from the Adoption of a Social Networking Tool." *Information Systems Research* 24 (1): 30–51.

Xu, Jennifer, Michael Chau, and Bernard CY Tan. 2014. "The Development of Social Capital in the Collaboration Network of Information Systems Scholars." *Journal of the Association for Information Systems* 15 (12): 835.

Yan, Lu (Lucy), Jianping Peng, and Yong Tan. 2015. "Network Dynamics: How Can We Find Patients Like Us?" *Information Systems Research* 26 (3): 496–512. https://doi.org/10.1287/isre.2015.0585.

Zhang, Xiaojun, and Viswanath Venkatesh. 2013. "Explaining Employee Job Performance: The Role of Online and Offline Workplace Communication Networks." *Mis Quarterly* 37 (3): 695–722.