

Communications of the Association for Information Systems

Volume 42

Article 16

4-2018

The Social Structure of the Information Systems Collaboration Network: Centers of Influence and Antecedents of Tie Formation

Wallace Chipidza

Baylor University, wallace_chipidza@baylor.edu

John F. Tripp

Baylor University

Follow this and additional works at: <http://aisel.aisnet.org/cais>

Recommended Citation

Chipidza, Wallace and Tripp, John F. (2018) "The Social Structure of the Information Systems Collaboration Network: Centers of Influence and Antecedents of Tie Formation," *Communications of the Association for Information Systems*: Vol. 42 , Article 16.

DOI: 10.17705/1CAIS.04216

Available at: <http://aisel.aisnet.org/cais/vol42/iss1/16>

This material is brought to you by the AIS Journals at AIS Electronic Library (AISeL). It has been accepted for inclusion in Communications of the Association for Information Systems by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.



The Social Structure of the Information Systems Collaboration Network: Centers of Influence and Antecedents of Tie Formation

Wallace Chipidza

Baylor University

wallace_chipidza@baylor.edu

John F. Tripp

Baylor University

john_trippp@baylor.edu

Abstract:

In this study, we examine the historical information systems research collaboration network. We build the network using coauthorship information in the Senior Scholars' basket of eight journals from the publication of *MISQ*'s first issue in April, 1977, to November, 2015. The different journals vary widely in their network configurations. We examine the influence of gender homophily, geographic homophily, and field tenure heterophily on coauthorship in the network. From using exponential random graph modeling (ERGM) on a randomly selected subset of the network, we present preliminary evidence that suggests that ties in the IS collaboration network exhibit homophily according to gender and geography. Conversely, coauthorship seems to exhibit heterophily along the temporal dimension: short-tenured researchers in the field prefer to collaborate with long-tenured researchers. ERGM enables one to make statistical inferences concerning the influence of node attributes and structural variables on network formation, which is hard to do with logistical regression because network relationships violate the independence of observations assumption. We also reveal the current center of the IS collaboration network. Based on this center, we propose a metric to measure a researcher's connectedness in the network.

Keywords: Social Network Analysis, Coauthorship, Collaboration, Homophily, Heterophily, Exponential Random Graph Modeling.

This manuscript underwent peer review. It was received 12/21/2016 and was with the authors for 9 months for 3 revisions. Fred Niederman served as Associate Editor.

1 Introduction

The information systems (IS) research field is one example of a community of practice in which knowledge results from social interaction (Gallivan & Ahuja, 2015). Researchers do not neutrally observe the research process; they exist in a temporal and sociological context, which implies that macro factors such as existing social norms and attitudes affect how researchers choose their collaborators, topics of study, and publication outlets. Uncovering the collaboration network of the IS field can enhance our understanding of how the field creates scientific knowledge and may reveal latent biases that could impact the trajectory of certain areas of research or impede the success of researchers associated with particular demographics.

Every academic field has a collaboration network that one can uncover by uncovering the links between all coauthors. One can view this network as an artifact of the field that can represent its history. Indeed, calls for recording and preserving the history of IS research have increased in recent years (Zhang, 2015). In this study, we analyze the historical IS collaboration network (CIS) and articulate its properties consistent with Mason, McKenney, and Copeland's (1997) articulation that "if MIS is to enjoy the theoretical and professional recognition that academic maturity bestows on a discipline..., MIS professionals must begin also to record and examine its history" (p. 258). Although Mason et al. (1997) focused mainly on the need to understand the history of organizations in the context of technological use, their argument equally applies to the IS research field as a whole.

Collaboration in a research field is positively associated with scholarly output, which forms the basis for reward assessment in academia (Gallivan & Ahuja, 2015). The connections that collaboration creates weave into a complex network in which researchers influence each other in the knowledge production process (Xu, Chau, & Tan, 2014). However, to date, little research has focused on understanding the dynamic nature of collaboration in IS research (Gallivan & Ahuja, 2015). As the volume of research in IS has grown over time, the field has accumulated enough structural information about its network to allow one to generate and analyze the dynamic collaboration network over time.

In line with past research that emphasizes the effects of social interactions on the knowledge production process in IS research, we investigate three research questions:

RQ1: How do author characteristics determine tie formation?

RQ2: What are the characteristics of the IS collaboration network?

RQ3: How do the subnetworks corresponding to the different journals differ if at all?

To answer RQ1, we build on past research to identify possible determinants of collaboration. Given that previous research largely focuses on homophily as a theoretical lens, we add the lens of heterophily—that is, preference for different others (Easley & Kleinberg, 2010)—as a theoretical basis to identify another determinant of such collaboration. Further, we examine whether these determinants have consistently shaped collaboration in IS over time or whether they have been more salient in certain eras compared to others, which is important because understanding the changes in collaboration preferences may have implications for the diversity of research topics that our top journals cover. As an example, if experienced researchers tend to collaborate only with other experienced researchers, then new research topics might not receive adequate attention. As such, the determinants of collaboration might ultimately implicate the novelty and relevance of research output.

To answer RQ2, we use a measure similar to the Erdős number—named after the Paul Erdős, the highly productive Hungarian mathematician who researchers have widely cited as the center of the math research network (Grossman, 2002). Mathematicians trace their connections to Paul Erdős through use of the Erdős number, a measure of the collaborative distance between any given author and Paul Erdős (De Castro & Grossman, 1999). To characterize the IS field, we need to capture characteristics of the network such as its centers of influence so we can similarly recognize the structure of influence in the network and those individuals who are most influential in the field. While scholarly influence is a multifaceted construct (Cuellar, Vidgen, Takeda, & Truex, 2016), we focus on one aspect of influence; namely, connectedness in the network of researchers. We further examine whether a researcher's centrality in the network as based on our selected journal list is positively related to research productivity in the broader IS field and beyond.

To answer RQ3, we examine variations in network configurations, which might suggest the field has asymmetrically accumulated social capital over time. For example, the different journals might focus on

understanding the configurations of their networks in order to decide whether they need to foster more collaboration.

To answer our research questions, we construct the IS collaboration network using coauthorship information for the period between 1977 (the year of *MIS Quarterly's* first issue) and 2015. We limit the analysis to the current “basket of eight” journals: *MIS Quarterly (MISQ)*, *Information Systems Research (ISR)*, *Journal of Management Information Systems (JMIS)*, *Journal of the Association for Information Systems (JAIS)*, *European Journal of Information Systems (EJIS)*, *Information Systems Journal (ISJ)*, *Journal of Strategic Information Systems (JSIS)*, and *Journal of Information Technology (JIT)*—a total of about 6,000 papers. We note that the proxy we use as the IS research network, the basket of eight journals, does not fully represent the field because other publication venues such as journals and academic conferences also advance knowledge in it. Nevertheless, we can garner important insights about the network from examining collaboration in these eight journals.

The paper proceeds as follows: in Section 2, we review related work and, in Section 3, outline the theory that underpins the study. In Section 4, we detail how we collected and analyzed our data. In Section 5, we present the results. In Section 6, we discuss our findings and conclude the paper.

2 Literature Review

Several studies have focused on structural aspects of the IS collaboration network (Xu et al., 2014; Zhai, Li, Yan, & Fan, 2014). Xu et al. (2014) examined coauthorship data from the basket of six journals for period between 1980 and 2012. They focused on describing the structural evolution of the network and found that, over time, the IS field has acquired social capital and become more connected. Further, the fraction (i.e., the number of papers that have at least two authors divided by the total number of published papers) and extent (i.e., the average number of authors per paper) of collaboration have also increased during that period. In fact, collaboration has increased to such an extent that it eclipses collaboration rates in other business fields such as management, marketing, and finance. The authors attributed this increase to the IS field's diversity in research topics and methods. Consequently, IS researchers form more collaboration ties per capita relative to other business fields, which leads to a more connected collaboration network. Network structure results from tie formation. We extend Xu et al.'s (2014) study by investigating the antecedents to tie formation. We explore the effects of author characteristics on tie formation in the IS network. In addition, by expanding the journal set to incorporate at least the basket of eight journals and the period after 2012, we should generate new information about the IS collaboration network.

Gallivan and Ahuja (2015) examined a set of hypotheses concerning coauthorship in five IS journals (*MISQ*, *ISR*, *EJIS*, *JMIS*, and *ISJ*). They also found that, over time, the fraction and extent of coauthorship in these journals had increased. They also found that the fraction and extent of coauthorship for journals published in North America (*MISQ*, *ISR*, and *JMIS*) was greater than for journals published in Europe (*EJIS* and *ISJ*). Furthermore, they found a positive relationship between the number of authors on a paper and its subsequent citations in three journals (*MISQ*, *JMIS*, and *EJIS*). The relationship was curvilinear with the inclusion of higher-order terms: as the number of coauthors per paper increased, the citations leveled off or decreased for the three journals. For *ISR*, the relationship between the number of coauthors and its citations was negative, and no such relationship existed for *ISJ*. These findings show that, although collaboration is beneficial, at a certain point the citation benefits of increasing the number of collaborators level off. Hence, collaboration in IS research is a complex phenomenon that warrants further examination.

The IS field has traditionally underrepresented women and minorities (Coder, Rosenbloom, Ash, & Dupont, 2009). This underrepresentation may extend to the IS academe as well. Gallivan and Ahuja (2015) examined whether IS researchers collaborate based on gender and institutional homophily—that is, preference for similar others (McPherson, Smith-Lovin, & Cook, 2001)—based on collaboration data from a combined 35 (seven chosen at random from each publication) volumes from *MISQ*, *JMIS*, *ISR*, *EJIS*, and *ISJ* from 1999 to 2005. They found that collaboration in these mainstream IS journals exhibited homophily according to gender and institution of graduation. The set of journals that Gallivan and Ahuja (2015) covered excluded three journals in the Senior Scholars' basket of eight journals and only covered the period between 1999 and 2005. Hence, the study drew its conclusions from a small subset of published research during a relatively short period; as such, further examining gender homophily could help strengthen the argument that such homophily generalizes to the entire period of the field's existence.

In that spirit, we expand the number of journals to eight and the period of examination from six years to 38 years. Further, in addition to gender homophily, we also examine the influence of geographic homophily and field tenure heterophily (the length of time since a researcher graduated with a PhD) on collaboration in IS research.

Another common theme in research network studies concerns metrics that capture a researcher's potential and/or past productivity. To that end, researchers have proposed various measures such as the h-index and the number of publications in selected journals (Lowry et al., 2013). Such studies typically find problems in the status quo; hence, they propose new metrics to assess scholarly influence. Most recently, Cuellar et al. (2016) proposed that, instead of counting a scholar's number of papers in certain journals, one should measure scholarly capital as an aggregation of three measures: ideation, venue representation, and connectedness. Venue representation refers to the publication outlets, such as journals and conferences, recognized in the field. Ideational influence refers to the uptake of a scholar's work, operationalized as an aggregation of the h (measures the number of times a researcher has been cited at least a certain number of times), g (weighs higher cited papers more heavily), and gc (weighs more recent papers more favorably) indices. Connectedness measures the extent to which one is connected to influential researchers in the field, operationalized as an aggregation of the author's degree, closeness, and betweenness centralities. Connectedness is an important measure of scholarly influence because it predicts a researcher's scholarly output (Lowry et al., 2013). In this study, we propose a complementary measure for connectedness—the Lyytinen number—that does not require full knowledge of the collaboration network to compute. One may view this measure as a proxy for betweenness; although it is less precise than betweenness centrality, the Lyytinen number is simpler to compute and may, therefore, be more useful to hiring and tenure promotion committees when making their decisions.

3 Theoretical Development

In the context of social networks, researchers have focused on the advantage or disadvantage that accrues to individual nodes (people) and to networks as a whole based on the structure of ties between nodes (Burt, 2002). Network ties do not form randomly. Instead, ties form because of social and other forces that create variation in the advantage a node may gain by sharing a tie with another node. While some network ties only exist because two nodes share a task or goal (e.g., ties formed by work assignment in a team), research collaboration ties do not generally follow this “mandatory” tie-formation paradigm. Previous research has posited that the similarity-attraction hypothesis—that interaction is more likely to occur among people with similar traits—as one force that drives the formation of voluntary network ties (Yuan & Gay, 2006).

To investigate RQ1, we use the theories of homophily and heterophily. Homophily is the tendency for people to form connections with others of similar backgrounds (Currarini, Jackson, & Pin, 2009). Consistent with the similarity-attraction hypothesis (which the idiom “birds of a feather flock together” reflects), ties in any real network are likely to form among people of similar characteristics (Lazarsfeld & Merton, 1954; McPherson et al., 2001). The self-categorization principle—the tendency for individuals to place themselves and others in categories according to characteristics such as gender, age, and race—also explains homophily (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). These categories allow individuals to categorize others as similar or dissimilar, which forms the basis for homophilous ties. One can explain homophily with information flow: people with similar characteristics are likely to communicate more easily than people with different characteristics (Egorov, Polborn, & Welcome, 2010). Research has shown the concept that homophily drives ties to be true in friendships and online social networks such as Facebook (Harris, 2013; La Fond & Neville, 2010).

3.1 The Effect of Gender on Collaboration

In IS collaboration, researchers will likely prefer coauthors of the same gender—a hypothesis that Gallivan and Ahuja (2015) have previously examined. However, they examined their hypothesis based on data from a period of only seven years (1999-2005); further examination would show whether this homophily generalizes over time. For example, male authors might find it easier to communicate with other male researchers and, hence, socialize more outside of the academic context. Female researchers may have common experiences such as overcoming the stereotype that technology-related fields are for males or struggling with balancing work and life (McIlwee & Robinson, 1992; Watts, 2009). With greater socialization, these researchers might have more opportunities to discuss nascent research ideas that potentially lead to collaboration. As these relationships grow, so do the opportunities for research ideas to

intersect. The greater socialization among similar researchers than dissimilar researchers constitutes an example of the social structuring of activities, where similar people are brought into contact more frequently than one would expect from chance (Feld, 1982). Thus, we hypothesize that:

H1a: Ties are more likely to form between people of the same gender than between people of different genders.

In addition to individual differences' driving the formation of ties, group differences may add additional variation. For instance, the extent to which women form homophilous ties is likely to differ from the extent to which men form homophilous ties in the IS field, which results from self-categorization: on average, individuals who belong to a majority group (in the IS field, men) places themselves in a larger category compared to individuals who belong to a minority group (in the IS field, women). As Currarini et al. (2009) show, larger groups tend to exhibit higher homophily in tie formation than small groups. The IS field has a low proportion of female IS researchers (Gallivan & Ahuja, 2015). To illustrate the hypothesized difference in homophily according to gender, consider an extreme case in which only one woman existed in the IS field. Because the field in such a case would have an extreme male majority, the woman would have no choice but to collaborate with men (assuming that all researchers preferred collaboration to sole authorship). The number of women in the field would have to continually increase in order for women to find it feasible to preferentially collaborate with other women at a certain point. Because actual estimates for the composition of female IS researchers range from 20 to 30 percent (Gallivan & Ahuja, 2015), we expect that women will demonstrate much less homophily than men in their choice of coauthors in the IS field. On the other hand, because men are a substantial majority in the IS field, men will likely demonstrate great homophily in their choice of coauthor(s). Thus, we hypothesize that:

H1b: Female researchers demonstrate less homophily in their coauthor choices than male researchers.

In addition, large groups tend to form more ties per capita than small groups (Currarini et al., 2009). If homophily is the primary mechanism through which collaboration occurs, then, according to the self-categorization principle, men have more potential collaborators in their category on average, as compared to women. Therefore, we expect that male IS researchers will, on average, collaborate with more people than female researchers. Thus, we hypothesize that:

H1c: Male researchers form more ties with other researchers on average than female researchers.

3.2 The Effect of Geography on Collaboration

The IS field likely exhibits homophily according to the geographic location of PhD-granting institutions: that is, geographical distance likely affects whether individuals form network ties (geographic homophily). PhD graduates are more likely to stay in the country they graduate in for a variety of reasons. Institutions in the same country are likely to have similar recruitment standards and expectations for tenure. Additionally, institutions are likely to have existing recruitment relationships with institutions located in the same geographical region. For example, many IS students that graduate from New York University find employment at the Massachusetts Institute of Technology and vice versa (Gallivan & Ahuja, 2015). Relocating to a different geographical region often costs more than staying in the same region. Researchers are more likely to collaborate with people they have relationships with, including researchers at their own institutions. Therefore, in addition to collaboration between students of the same PhD program (Gallivan & Ahuja, 2015), we have also captured collaboration between faculty members and their own PhD students. For this study, we define a geographical region as a continent (Africa, Asia, Australasia, Europe, North America, and South America). Researchers in the same geographic region are also likely to have more opportunities for contact than one would expect from chance. Such opportunities come in the form of regional workshops, research colloquia, and industry and academic conferences. When researchers meet at these venues, they have opportunity to exchange and discuss research ideas, which may lead to collaboration. Thus, we hypothesize that:

H2: Ties are more likely to form among researchers that graduated in the same geographical region than among researchers from different geographical regions.

3.3 The Effect of Field Tenure on Collaboration

However, we have reason to believe that, in academic research, individuals may intentionally seek to form ties with people with different characteristics. When ties form between nodes with different properties at a greater level than one would expect from chance, the network exhibits heterophily (Currarini et al., 2009). Academic researchers may elect to collaborate with people that possess core competencies that they lack or have published previously in a particular journal. If enough authors exhibit a tendency to collaborate based on resource complementarity in the network, then heterophily might explain tie formation in the network. The desire to complement resources might then dwarf other factors that potentially affect collaboration in the network, such as gender, race, or geographic location. This tension between forces that drive homophily and forces that drive heterophily makes the formation of network ties a complex phenomenon to research.

The IS field also likely exhibits heterophily according to field tenure (temporal heterophily). For this study, we define field tenure as the length of time since researchers received their PhD. Researchers would seek to collaborate with authors who possess skills and qualities that they lack (Cuellar et al., 2016), which grants them advantages that may strengthen their chances that journals will publish their papers. Senior faculty often have substantial experience in publishing in elite journals such as the basket of eight. Junior researchers, on the other hand, likely lack this experience. Hence, in seeking out collaborators, junior researchers, such as PhD students and junior faculty, are likely to prefer collaborators with greater experience. In the same vein, senior researchers are also likely to prefer collaborating with junior researchers with whom they have mentoring relationships. Thus, we hypothesize that:

H3a: Ties are more likely to form between short-tenured researchers and long-tenured researchers than one would expect from chance.

Furthermore, the more time researchers are part of the field, the greater the human capital that they acquire. This capital comes in the form of research ideas and human connections. The longer that researchers are part of the field, the more research ideas they might produce, which make them more attractive as collaborators. Moreover, a longer-tenured researcher would have more opportunities for contact with other researchers than short-tenured researchers. As a result, a researcher with an older PhD will have more opportunities to collaborate than a researcher with a recent PhD. Thus, we hypothesize that:

H3b: Long-tenured researchers are likely to have more ties with other researchers than short-tenured researchers.

Table 1 below summarizes the hypotheses.

Table 1. Research Hypotheses

	Hypothesis
H1a	Ties are more likely to form between people of the same gender than between people of different genders.
H1b	Female researchers demonstrate less homophily in their coauthor choices than male researchers.
H1c	Male researchers form more ties with other researchers on average than female researchers.
H2	Ties are more likely to form among researchers that graduated in the same geographical region than among researchers from different geographical regions.
H3a	Ties are more likely to form between short-tenured researchers and long-tenured researchers than one would expect from chance.
H3b	Long-tenured researchers are likely to have more ties with other researchers than short-tenured researchers.

4 Method

To create the collaboration network, we used a subset of coauthorship ties in the IS field; namely, those ties in the Senior Scholars' basket of eight journals. The Senior Scholars' basket of eight¹ journals is a widely accepted group of IS journals that scientometric studies and literature reviews frequently employ as a representative journal collection (Li & Karahanna, 2015; Zhai et al., 2014). As such, we selected the basket of eight journals as our sample of the IS field collaboration network. We believe the basket of eight suits our study for two reasons. First, the basket is a widely used standard for evaluating tenure and promotion cases, which means that these journals provide researchers with major incentives to publish in them (Lowry et al., 2013). Because at least a subset of the journals within the basket of eight are universally accepted "A"-level publications, they draw a widely dispersed set of contributions, and we believe that focusing on these journals allows one to investigate a very large segment of the connections in the IS field. Second, from a pragmatic standpoint, we selected this subset of the IS research field to keep the data collection effort manageable because a large portion of the data collection required manual effort. However, our sample is significantly larger than those samples that previously published papers on the IS collaboration network have used, and we believe that it is a sufficiently representative sample to investigate the research questions that we pose². Appendix A discusses the technical details of our data-collection and parsing process.

To investigate RQ1, we used exponential random graph modeling (Hunter, Handcock, Butts, Goodreau, & Morris, 2008), which we explore in Section 4.1. For H3a, for which we had to examine a smaller subset of the sample edges, we calculated the heterophily by hand using the test for homophily/heterophily that Easley and Kleinberg (2010) explicate.

To investigate RQ2, we constructed the combined network using R's igraph library (Csardi & Nepusz, 2006; Ihaka & Gentleman, 1996). Each unique author in the network constituted a node, and we used each author's first and last name as the node's unique identifier. We then obtained descriptive characteristics of the network, particularly the number of nodes, number of edges, average shortest path, and degree distribution, using igraph's inbuilt network functions. We also calculated betweenness and eigenvector centrality scores for each node in the network in order to identify the most connected researchers in the IS field. To investigate RQ3, we repeated all the steps needed to answer RQ2. By doing so, we could compare the characteristics of each journal, particularly each journal's size and connectedness. We also visualized the different subnetworks with igraph's "plot" function, so we could visually examine the network configuration differences among journals.

4.1 Exponential Random Graph Modeling

Until recently, most work in social network analysis has focused on describing certain quantifiable measures of a network. For example, researchers have generally used transitivity as a proxy for network connectedness and posited that higher levels of transitivity signal better connectedness (Wyatt, Choudhury, & Bilmes, 2008). In the same vein, researchers have measured the extent to which nodes mix according to some criterion using the assortativity measure (Handcock, Hunter, Butts, Goodreau, & Morris, 2008). While these measures are useful in describing a network, they do not explain or predict how a network forms.

Exponential random graph modeling (ERGM) is useful because it helps one to explain or predict the probability that a network tie exists between two nodes based on some attributes of the nodes (Handcock et al., 2008). ERGM enables one to test the statistical significance of certain tendencies in the network. The simplest of these tests is whether a network displays a tendency to form edges. Real-world networks are usually sparsely populated relative to the maximum possible number of edges. Hence, an ERGM test for the tendency of a network to form ties will usually display a negative coefficient. A more interesting example is whether a network demonstrates a tendency toward transitivity. With traditional social network analysis, one would calculate the clustering coefficient of the network and, if it is sufficiently high, conclude that the network has transitivity. But what exactly counts as a high clustering coefficient? ERGM answers that question through comparing the coefficient with that of a randomly generated network of the same number of nodes and number of edges. More sophisticated tests are also possible; for example, one can

¹ "Senior Scholars' Basket of Journals," Association for Information Systems, <https://aisnet.org/?SeniorScholarBasket>.

² This is not to say that research published in the basket of eight is the only valid IS research, as there are certainly many journals that publish very highly cited IS research, in addition to many excellent academic conferences.

test whether the network displays a tendency for 1) nodes to be isolated, 2) nodes to reciprocate ties (in directed networks), and/or 3) homophily according to some node characteristic. Therefore, ERGM enables the researcher to model a network and, hence, provide a theory for explaining and predicting (Gregor, 2006) using network analysis. Specifically, we employ ERGM to predict the probability of a tie between any two nodes according to node characteristics.

In this study, we use ERGM to examine what effects specific researcher attributes—that is, gender, geographic region of PhD training, and field tenure—have on the probability that any two researchers will collaborate. By using ERGM, we could test the statistical significance of the network's tendency to display gender and geographic homophily.

ERGM is analogous to a binary logistic regression modeling in that the outcome variable is binary: a network tie exists between any pair of nodes or it does not (Wasserman & Pattison, 1996). The ERGM algorithm considers the social network as a set of random variables and assigns each variable in the set a probability (Wyatt et al., 2008). ERGM allows one to assign probabilities to a tie between any two nodes based on some property. The following equation describes the model:

$$p(Y = y|\theta) = \frac{1}{Z} e^{\theta^T \phi(y)},$$

where Y is the set of variables that represent edges in the network, y is the observed network adjacency matrix, Z is a normalizing constant that ensures that the probabilities stay within the 0 to 1 range and the probabilities across all networks sum up to 1, θ represents a vector of weights to be learned, and ϕ represents the feature functions defined on y (Harris, 2013; Wyatt et al., 2008). The equation above means that the model examines a set of random networks and assigns each of them probabilities that sum up to 1. Hence, the conditional probability of a random network Y , given the observed network y , depends on the exponent of a weighted set of statistics $\phi(y)$. The set of statistics represents the predictors that one can put in the model; for example, the number of edges, the number of triangles, and a gender homophily effect in the model. θ is a set of coefficients for the set of predictors.

To examine our hypotheses, we employed the ERGM package in R (Hunter et al., 2008). Undergirding the package's computation is a Markov Chain Monte Carlo (MCMC) algorithm that simulates the random sequence of networks. The sequence is a Markov chain that depends on only the current state of the simulated network. The algorithm accepts the next network conditionally based on the result of comparing it to the current one. Social network research widely uses the package, and research, which includes papers in *MISQ* and *Organization Science*, have cited it at least 390 times since 2008 according to Google Scholar³ (Chen, Chiang, & Storey, 2012; Faraj, Jarvenpaa, & Majchrzak, 2011; Goh, Gao, & Agarwal, 2016).

Large networks are usually not feasible to examine or visualize in their entirety. Thus, a sampling strategy that selects a representative subnetwork of the network can be useful in demonstrating aspects of the network. The collaboration network for IS research, C_{IS} , has 5670 nodes and 10303 edges, and it is not particularly large compared to other real-life networks such as older academic disciplines, Facebook, and the Internet (Barabási & Bonabeau, 2003). However, we can only collect collaboration information with an automated process from journals' tables of contents; one must manually collect demographic information such as gender and year of graduation for each author. Because the effort required to manually collect the information is huge, we found it prudent to select a subsample of the network on which to perform our analyses. A variety of sampling strategies exist; these strategies include random node sampling, random edge sampling (RES), random walk sampling, and random node neighbor sampling (Leskovec & Faloutsos, 2006).

For this study, we employed random edge sampling the structure of the resulting subnetwork is well understood (Leskovec & Faloutsos, 2006). The main drawbacks of random edge sampling are that the resulting subnetwork will be biased toward high-degree nodes and the resulting subnetwork will be sparsely connected and, thus, not respect the underlying community structure (RES is still better at respecting network structure than random node sampling) (Leskovec & Faloutsos, 2006). The former drawback arises because, by definition, high-degree nodes have many more edges per node than the average node. As a result, a high-degree node has quite a high probability of being part of a randomly selected edge list. C_{IS} has a power law distribution, and new members join the network through

³ See https://scholar.google.com/scholar?cites=6039954908351403189&as_sdt=5,44&scioldt=0,44&hl=en

preferential attachment (Xu et al., 2014). Therefore, understanding the homophily in an RES subnetwork can help one reveal the behavior of the most popular nodes in the network. The main advantage of RES sampling is that it is relatively simple. To conduct ERGM on C_{IS} , we randomly selected 110 ties from the consolidated edge list of C_{IS} . We carried out Google searches to discover demographic information for the randomly selected 208 researchers. For each researcher, we determined their gender from their first names and confirmed the gender by visiting their personal or institutional websites for photographs and biographical information. We also obtained other information by looking at publicly available researcher data from their CVs and websites such as LinkedIn and ResearchGate. We collected information about gender, year the researchers received their PhD (where applicable), and geographical location of PhD awarding institution. The above three attributes are fixed and, hence, the simplest attributes on which to anchor the ERGM.

5 Results

In presenting our results, we distinguish between descriptive and modeling statistics. Descriptive statistics, particularly the betweenness measure, are useful in revealing the current center of C_{IS} . Modeling statistics reveal how homophily and heterophily influence edge formation in the network.

5.1 Descriptive Statistics

Up until December, 2015, the basket of eight journals had published a total of about 5,500 research papers (excluding book reviews and editorials). *JMIS* had the highest number of publications and *ISJ* the lowest. About 22 percent had only one author, and 78 percent had at least two authors, which suggests that IS researchers prefer collaboration over sole authorship. Overall, *JAIS* had the highest number of authors per research paper and *EJIS* the lowest (see Table 2). Compared to the overall mean number of authors per paper (2.32), four journals had a higher statistic. Interestingly, three of these four journals also form the top tier of journals in the rankings that the AIS provides⁴.

Table 2. Collaboration Statistics for Basket of Eight Journals

	Number of publications	Single author papers	Multiple author papers	Total number of authors	Number of authors per paper
<i>MISQ</i>	1087	211	876	2538	2.33
<i>ISR</i>	671	89	582	1703	2.54
<i>JMIS</i>	1154	183	971	2889	2.50
<i>JAIS</i>	410	63	347	1077	2.63
<i>ISJ</i>	378	107	271	815	2.16
<i>EJIS</i>	812	281	531	1671	2.06
<i>JIT</i>	552	152	400	1156	2.09
<i>JSIS</i>	499	165	334	1045	2.09
Combined IS	5563	1,251	4312	12894	2.32

On average, each author in the IS research network collaborated with 4.19 authors with a standard deviation of 6.79. Figure 1 shows the degree distribution of C_{IS} , compared to the degree distribution of a random network of the same size and edge density. The distribution of node degrees followed a power law distribution; hence, C_{IS} resembles other real-world networks. The power law distribution of node degree means that there are few highly popular nodes and a majority of less popular nodes. A randomly generated network, on the other hand, would have a normal distribution of ties per node. The power law degree distribution implies that C_{IS} was scale free—the network grew with time; as such, new members had opportunities to join the network. Further, the network grew with preferential attachment: new members joined the network by collaborating with more established researchers. As a result, already popular nodes were more likely to acquire more links and, hence, the structure of the network reflects the “rich get richer” model of tie formation (Barabási & Bonabeau, 2003).

⁴ See <https://aisnet.org/general/custom.asp?page=JournalRankings>

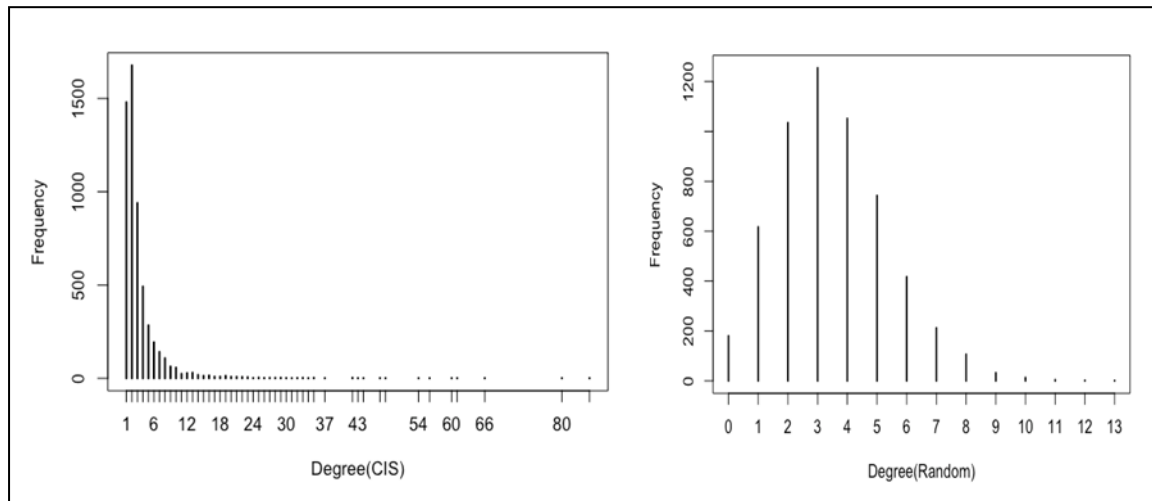


Figure 1. Degree Distributions of C_{IS} vs. Randomly Generated Graph of Comparable Size and Edge Density

Table 3 shows the different network metrics for the journals in the basket of eight. *MISQ*, *EJIS*, *ISR*, and *JMIS* had a high number of authors. *MISQ* had the largest connected component with 781 vertices. It also had the longest diameter (i.e., 17)—the maximum shortest path from one author to another. On average, in the *MISQ* network's largest component, an author had to take about seven steps to find another author. The diameter for C_{IS} was 20, and the average shortest path length was 7.38.

Table 3. Network Metrics for Journals in the Basket of Eight

Journal	Number of nodes	Number of edges	Proportion of largest component	Diameter	Average path length
<i>MISQ</i>	1479	2178	0.52	17	6.66
<i>ISR</i>	967	1515	0.56	15	5.72
<i>JMIS</i>	1607	2475	0.47	13	4.48
<i>JAIS</i>	786	1091	0.24	9	3.19
<i>EJIS</i>	1052	1244	0.15	6	1.85
<i>ISJ</i>	613	779	0.06	4	1.55
<i>JSIS</i>	739	805	0.05	7	2.59
<i>JIT</i>	933	991	0.14	7	2.85
Combined IS	5670	10303	0.67	20	7.38

Figure 2 visualizes the collaboration network for each different journal. *MISQ*, *ISR*, and *JMIS* had relatively large, discernible cores. Because *MISQ*, *ISR*, and *JMIS* consistently occupy high ranks on business journal rankings (Lowry et al., 2013), a stable core seems to be a requirement for (or perhaps a consequence of) journal success. It is also possible that senior researchers that have previously published in these journals have continued to do so with new students every few years, which has made the core larger and more densely connected. In other words, even though *MISQ*, *ISR*, and *JMIS* had large cores, researchers may have found it difficult to enter the network unless they collaborated with someone with existing publishing experience in these journals. On the other hand, *ISJ* and *JSIS* had sparsely populated networks and very small (if not non-existent) cores. In addition, C_{IS} had an even greater core that comprised 67 percent of the network—an increase from the 65 percent that Xu et al. (2014) report in describing the 1980-2012 IS network. This increase strongly supports the argument that C_{IS} has become more connected with time.

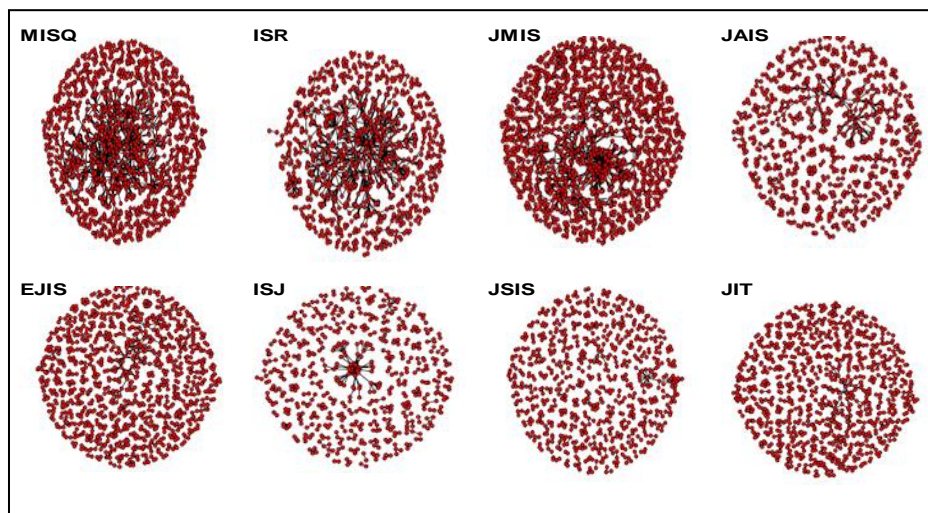


Figure 2. Network Configurations for the Basket of Eight Journals

C_{IS} , the combined network of all the journal networks, was more connected than any of the individual networks. It had a dense, stable core (Figure 3), which means that the network was robust and could survive the loss of its key nodes. Such existential debates as have dominated the discourse in the past perhaps lack as much relevance now. The academic community has increasingly come to recognize IS as a legitimate field that requires little justification for its existence. The robustness of C_{IS} is an empirical manifestation of its legitimacy because it shows that the field has acquired substantial social capital since its inception. Alternatively, the stable core may evidence that the same authors have repeatedly published in the basket of eight and, thus, potentially excluded those without prior publishing history unless they have collaborated with the established authors in these journals.

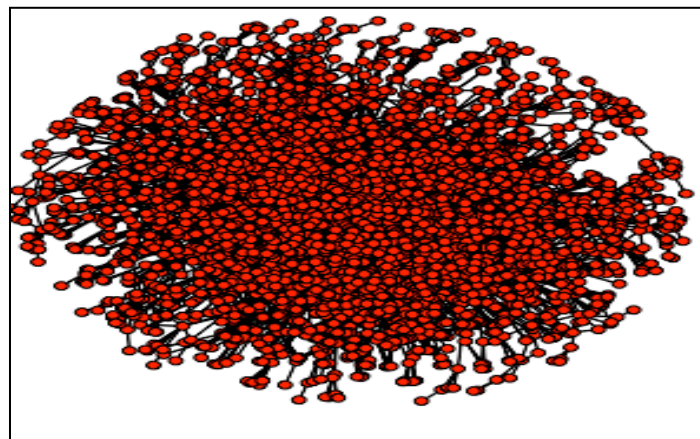


Figure 3. Network Configuration for C_{IS}

Table 4 shows the ten most central IS researchers according to betweenness centrality—a measure of the extent to which a network node lies between other nodes (Xu et al., 2014). Researchers with high betweenness scores act as brokers of knowledge transfer in the network (Xu et al., 2014). As such, these researchers may exercise a high degree of influence over the direction of the field because they link together researchers that are far apart in the network. Therefore, highly connected researchers provide the social mechanism that can bring researchers interested in diverse topics together. The majority of the most central IS researchers have served as editors in chief of the field's elite journals. For example, Izak Benbasat and Ritu Agarwal served as editors in chief for *ISR*, Kalle Lyytinen served as editor in chief for *JAIS*, and Arun Rai is the current editor in chief for *MISQ*, a post that Detmar Straub has occupied in the past; hence, evidence suggests that betweenness centrality is valuable and informative. We emphasize that betweenness centrality does not measure scholarly impact or productivity; rather, it measures how connected a researcher is in the network. Some highly productive researchers may prefer sole authorship to collaboration or they may publish in other prestigious journals not in the basket of eight (e.g., *Management Science* or *Academy of Management Journal*); hence, productive scholars will not

necessarily have high centrality scores in C_{IS} . Nevertheless, as we show below, highly productive researchers were likely to have higher than average betweenness scores.

Table 4. Highly Connected IS Researchers According to Betweenness Centrality

Name	Current affiliated institution	Betweenness centrality
Kalle Lyytinen	Case Western Reserve (USA)	597033.45
Izak Benbasat	University of British Columbia (Canada)	519542.69
Alan Dennis	Indiana University (USA)	453721.58
Paul Pavlou	Temple (USA)	407611.29
Ritu Agarwal	University of Maryland (USA)	388043.36
Detmar Straub	Temple University (USA)	338316.95
Arun Rai	Georgia State (USA)	326785.77
Gordon Davis	University of Minnesota (USA)	309470.54
Jay Nunamaker	University of Arizona (USA)	298837.53

Kalle Lyytinen of Case Western Reserve University had the highest betweenness centrality score in C_{IS} , which means that his removal from the network would penalize everyone else's average shortest path length the highest. By the betweenness measure, Lyytinen qualifies as the current center of the historical C_{IS} and, is, therefore the IS equivalent of Paul Erdős. Lyytinen has published prolifically both in North American and European journals. As we describe in Section 1, in math research, the Erdős number illustrates the collaboration distance between an author and Paul Erdős. Similarly, we define the Lyytinen number (LN) of an author in IS research as the collaboration distance between an author and Kalle Lyytinen.

An author that has collaborated with Lyytinen has an LN of 1; an author that has not directly collaborated with Lyytinen but has collaborated with a coauthor of Lyytinen has an LN of 2, and so on. The minimum LN is 0, which corresponds to Lyytinen's collaboration distance from himself. The maximum Lyytinen number in the largest component of C_{IS} was 10, which was also Lyytinen's eccentricity. As we state above, the largest component of C_{IS} contained 67 percent of all researchers; hence, 67 percent of all researchers in C_{IS} had an LN. The median LN in the largest component of C_{IS} was 4 (mean = 3.84, SD = 1.33). For the most productive researchers in the IS field (see footnote below), the median LN was 3 (mean = 2.85, SD = 1.14). Figure 4 conveys this information. The differences between the two group means was statistically significant ($F = 68.53$, $df = 1$, $p = 0.000$). These statistics suggest that top IS researchers have low LNs. The same phenomenon is evident in math research where successful mathematicians tend to have lower Erdős numbers than the average (De Castro & Grossman, 1999).

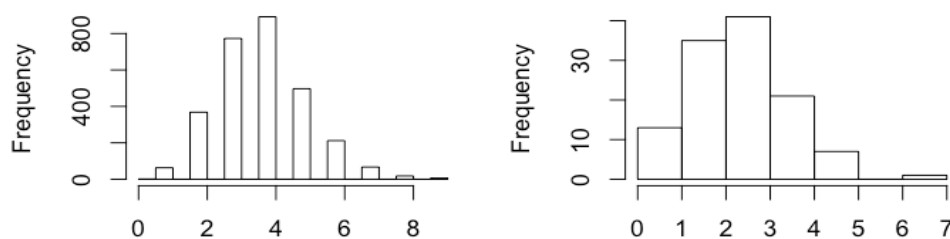


Figure 4. Distribution of Lyytinen Numbers for Authors in C_{IS} (Left Histogram) and Top IS Researchers (Right Histogram)⁵

Alternative measures of network influence also exist, such as degree, closeness, and eigenvector centrality (Bonacich, 2007). The latter measure is particularly useful because it captures the importance of a node's connections: researchers with high eigenvector centrality are well connected with other important researchers (Bonacich, 2007). Table 5 shows different journal combinations and the researcher with the highest betweenness and eigenvector centrality score across each combination. From the table, one can convincingly argue that Izak Benbasat and Jay Nunamaker consistently occupy the most influential

⁵ We employed the H-Index list maintained at the University of Arizona (see <https://ai.arizona.edu/sites/ai/files/MIS510/h-index-2015-04.pdf>).

positions in the historical IS research network. However, in identifying the center, we chose the betweenness measure because it measures the extent to which researchers in a network view a researcher as a broker of knowledge transfer (Xu et al., 2014). Scholars with high betweenness scores can act as links between other scholars from different parts of the network (Cuellar et al., 2016). Table B1 in Appendix B shows the five most connected researchers in each journal according to betweenness and eigenvector centralities. Over time, the center has changed. Before 1989, Jay Nunamaker and Benn Konsynski were the most connected according to betweenness and eigenvector centrality, respectively. In the 1990s, Joseph Valacich and Jay Nunamaker were the most connected. From 2000 to 2015, the most connected researchers according to betweenness and eigenvector centralities were Kalle Lyytinen and Jay Nunamaker, respectively.

Table 5. Highly Connected IS Researchers across Journal Combinations

Journal combination	Highest betweenness centrality score	Highest eigenvector centrality score
Basket of eight	Kalle Lyytinen	Jay Nunamaker
MISQ + ISR	Izak Benbasat	Detmar Straub
MISQ + ISR + JMIS	Izak Benbasat	Jay Nunamaker
MISQ + ISR + JMIS + JAIS	Izak Benbasat	Jay Nunamaker
Basket of six	Izak Benbasat	Jay Nunamaker

5.2 Modeling Statistics

The null model represents the baseline model that we can compare our model against after accounting for the influence of node attributes on the formation of a tie in the model (Harris, 2013). Akaike’s “an information criterion” (AIC) is a formula that one can use to compare models that are fitted according to maximum likelihood to the same data (Akaike, 1974). The smaller the AIC, the better the model. The following R code builds the baseline model:

```
> ergm( nrelations ~ edges)
```

The edges term demonstrates the propensity for ties to form in the network, which is typically low in real-world networks (Harris, 2013); we can conclude from the negative coefficient that the structure of the network features a low probability of edge formation.

Table 6. Null Model for RES C_{IS} Subnetwork (AIC = 1382)

	Coefficient	Std. error	p-value
edges	-5.27151	0.09558	.000***

Having created the baseline model, we examined the effect of node attributes on tie formation in the network. We tested the homophily effects in the model to test hypotheses H1a and H2. To do so, we had to add the gender, geographical region of PhD program, and period of PhD graduation researcher attributes to the model as main effects. We present the null and alternative hypotheses for the main effects of researcher gender, time of graduation, and geographic region of PhD training on the likelihood of coauthorship below:

- H₀ (gender homophily):** For any two researchers, no relationship between their genders and the likelihood that they will collaborate exists.
- H_A (gender homophily):** For any two researchers, a relationship between their genders and the likelihood that they will collaborate exists.
- H₀ (temporal homophily):** For any two researchers, no relationship between the times that they attained their PhDs and the likelihood that they will collaborate exists.
- H_A (temporal homophily):** For any two researchers, a relationship between the times that they attained their PhDs and the likelihood that they will collaborate exists.
- H₀ (geographic homophily):** For any two researchers, no relationship between the geographical regions in which they attained their PhDs and the likelihood that they will collaborate exists.

H_A (geographic homophily): For any two researchers, a relationship between the geographic regions that they attained their PhDs and the likelihood that they will collaborate exists.

We use the edges term, gender, geographic, and temporal homophily terms as predictors of the probability that a tie would form. The ERGM command initiates an MCMC algorithm that estimates the probability. The homophily terms capture the degree to which nodes of similar characteristics tend to form edges over or below what one would expect from chance. The “nodematch” command tests for homophily. The results in Table 7 show positive and significant homophily coefficients for gender and geographical region of PhD program but not for the period of PhD attainment in the subnetwork of 208 authors and 110 edges. Therefore, we reject the null hypotheses for the gender and geographic attributes, but we fail to reject the null hypothesis for the temporal attribute. In addition, the AIC decreased after adding the homophily terms (compared to the AIC for the null model), which indicates that the homophily model fits the observed data better than the null model. Hence, we found empirical support for H1a and H2. The following R code builds the model that tests for gender, geographic, and temporal homophily in the network.

```
> ergm( nrelations ~ edges + nodematch('Gender') + nodematch('Region') + nodematch('GradPeriod'))
```

Table 7. Main Effects Model for RES C_{IS} Subnetwork (AIC = 1356)

	Coefficient	Std Error	P-value
Edges	-6.3709	0.2728	.000***
Gender	0.4577	0.2281	.0448*
Region	1.0878	0.2279	.000***
GradPeriod	0.187	0.2149	.3842

Note that we conducted journal-by-journal analysis of homophily influences in the network and found no significant homophily influences in any of the journals. As we collect more data, such influences will likely show up if they exist.

We ran a multiple regression test to determine the effect of an author's gender and tenure of PhD on that author's degree centrality. The degree centrality of authors varied according to the tenure of their PhDs ($\beta = 0.41$, $p = .000$) but not according to an author's gender. This difference was not significant ($p = .215$). Figure 5 shows kernel density plots that correspond to the degree distributions for male versus female researchers, and it shows similar distributions between the two groups. The long-tailed degree distributions are consistent with the preferential attachment mechanism prevalent in real-world networks in which most nodes have low degrees and a handful of nodes have high degrees (Barabási & Bonabeau, 2003). The tenure of the PhD explained 10 percent of the variance in degree centrality. Specifically, the longer the tenure of the PhD, the greater the number of collaborators. These findings provide support for H3b but not for H1c.

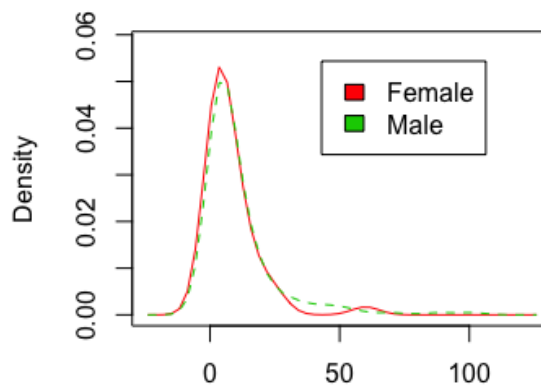


Figure 5. Number of Collaborators by Gender

We ran the *mixingmatrix* command in R so that we could visualize the relative proportions of MM versus FF versus FM or MF edges in the RES subnetwork. Table 8 shows that in the sampled subnetwork, MM edges comprised the majority of coauthorship links (70%) followed by heterophilous links (23%) and FF

links (7%). FM links are the exact same as MF links; hence, the total number of links was 77+25+8. The expression $2p(1-p)$ provides the expected number of heterophilous links, where p is the proportion of males and $1-p$ is the proportion of female nodes (Easley & Kleinberg, 2010). We compared the expected number with the observed proportion of female-female links relative to all the links they were involved in (25/33). We conducted a one-tailed t-test and found a significant difference between the two ($t = 3.718$, $df = 32$, $p\text{-value} = 0.000$). This finding lends support to H1b. The following R code displays the counts of co-authorship ties cross-tabulated by gender.

```
> mixingmatrix(nrelations, 'Gender')
```

Table 8. Collaboration Combinations According to Gender

Right-side author Left-side author	Male	Female
Male	77	25
Female	25	8

To investigate H3a (i.e., that ties are more likely to form between short-tenured researchers and long-tenured researchers than one would expect from chance), we selected only a subset of the sample of the edges because we found it difficult to identify a time that evidenced a researcher as junior given the likely variation in tenure-acquisition periods across time. Hence, to simplify the analysis, we classified researchers as junior if and only if they acquired their PhDs after 2009. By doing so, we could be reasonably confident that the researcher would have authored the paper as a junior researcher. To investigate the hypothesis that junior researchers select coauthors based on heterophily, we selected only those edges that included a junior author in them: a total 24 edges. We compared the expected number of heterophilous edges based on the junior/senior researcher composition against the observed number of heterophilous edges using a one-tailed test of significance. The difference was significant at the 0.05 level ($t = 3.7006$, $df = 23$, $p\text{-value} = 0.0005898$). We performed a robustness check: the hypothesis also held if we relaxed the definition of a junior researcher as one that acquired a PhD after 2005.

Table 9 summarizes our findings.

Table 9. Research Outcomes

	Hypothesis	Result
H1a	Ties are more likely to form between people of the same gender than between people of different genders.	Supported
H1b	Female researchers demonstrate less homophily in their coauthor choices than male researchers.	Supported
H1c	Male researchers form more ties with other researchers on average than female researchers.	Not supported
H2	Ties are more likely to form among researchers that graduated in the same geographical region than among researchers from different geographical regions.	Supported
H3a	Ties are more likely to form between short-tenured researchers and long-tenured researchers than one would expect from chance.	Supported
H3b	Long-tenured researchers are likely to have more ties with other researchers than short-tenured researchers.	Supported

We investigated whether homophily shaped collaboration consistently over the different eras. We found that geographic homophily was consistently influential across three different eras: pre-1989, 1989-2000, and 2000-2015. We observed gender homophily only after the year 2000. Field tenure homophily was not influential across the different time periods. Table 10 summarizes these findings.

Table 10. Homophily Influences Across Time

Era	Gender homophily	Field homophily	tenure	Geographic homophily
1977-1988	No	No		Yes
1989-2000	No	No		Yes
2000-2015	Yes	No		Yes

Finally, we calculated odds ratios for the homophily effects in the RES network. The odds ratios in Table 10 show that two researchers of the same gender were 1.58 times more likely to collaborate than two researchers of different genders (95% CI: 1.01 to 2.47). Further, two researchers that graduated from the same geographical region were three times more likely to collaborate than two researchers that graduated from different geographical regions (95% CI: 1.9 to 4.64). On the other hand, two researchers that graduated at around the same time were *not* any more likely to collaborate than two researchers that graduated in different eras.

Table 11. Odds Ratios for Homophily Model

	Lower	Odds ratio	Upper
Gender	1.01	1.58	2.47
Region	1.9	2.97	4.64
GradPeriod	0.79	1.21	1.84

6 Discussion

This study makes several contributions toward our understanding of collaboration in the IS research field. First, we assess the role of homophily in determining collaboration in the field. Past research has examined gender and geographic homophily (Gallivan & Ahuja, 2015). According to the self-categorization principle, researchers categorize themselves across different measures. Hence, we assessed geographic homophily differently than in previous studies. Whereas Gallivan and Ahuja (2015) define geographic homophily as attending the same PhD program, we broaden the definition to refer to the region of the PhD program. The difference between the two measures is subtle, but it allows one to capture collaborations between faculty and their (former) students (see explanation in Section 3); Gallivan and Ahuja's definition of geographic homophily would not capture this collaboration. As such, with this definition, we could examine the prevalence of cross-continental collaboration in IS versus intracontinental collaboration.

We found support for gender and geographic homophily. With the rising prominence of cloud-enabled research collaboration tools such as Dropbox and Zotero, it is possible that geographic homophily will disappear with time because authors will not need to meet in person when coauthoring research papers. Individual researchers might reflect on whether they should consider collaboration with researchers based in other locations, especially when they meet at conferences and workshops. More heterophilous research according to gender and geography may result in higher levels of idea sharing and, thus, in higher-quality research and an increased focus on understudied areas of the field. For example, the topic of ICT4D has received limited focus in the basket of eight journals (Venkatesh, Bala, & Sambamurthy, 2016; Walsham & Sahay, 2006) possibly because has little representation in the IS field.

In addition, we also examine how the effects of homophily have changed over time. In the period before 1989, gender homophily effects were not apparent because the proportion of males in the field was very high (about 95%). Remember that the Schelling model of homophily compares the expected number of heterophilous links $2p*(1-p)$ with the observed number of heterophilous links. As p , the proportion of males in the field, approaches 100 percent, the expected number of heterophilous links approaches zero—as does the observed number of heterophilous links. Hence, we unsurprisingly did not observe homophily before 1988. This proportion dropped to 88 percent in the succeeding decade and to 75 percent in the 2000-2015 period. Hence, the proportion of women has gradually increased over the years. Thus, from 1990 to 2015, gender homophily has become more pronounced. To the best of our knowledge, we are the first to examine whether the effects of homophily have been constant over different periods of time.

We also examine the role of heterophily in determining collaboration in the IS field. We found that junior researchers preferred to collaborate with senior researchers more than one would expect from chance. This finding makes sense given the preferential attachment mechanism that shapes the network. Because senior researchers can complement what junior researchers lack in experience and expertise, the latter can benefit greatly from collaborating with their seniors. Heterophily also explains how female researchers chose their collaborators. However, this heterophily might be a function of relative group sizes. The low proportion (21%) of female researchers in the historical network presents difficulties in finding other female collaborators. The contribution to research lies in showing that collaboration in the IS field is a product not only of homophily but also heterophily when one considers different groups. Future research could examine whether this differential homophily exists in other demographic categories such as race

and country of origin. It could also examine homophily according to research competencies (e.g., whether researchers well versed in theory select researchers that are competent in certain methods). To our best knowledge, we are the first to examine the role of heterophily in shaping collaboration in the IS field.

Our study differs from previous collaboration research in two different ways. We examine collaboration across all journals in the basket of eight journals, and our analysis covered from 1977 to 2015. By analyzing such a period, we could more holistically examine the historical IS network.

Although multiple studies recognize the importance of a researcher's connectedness in the research network (Cuellar et al., 2016; Lowry et al., 2013), none actually provide an easily accessible measure. Cuellar et al., (2016) propose that one should measure connectedness as an aggregation of degree, closeness, and betweenness centralities. To calculate these centralities, one would need to construct the whole research network, perhaps from the basket of eight journals and IS conferences, which is difficult to accomplish. Our revelation of Kalle Lyytinen as the current center of C_{IS} provides an easy metric of approximating a researcher's embeddedness in the network. The LN for a researcher is a proxy for their distance from the center of the research network. For any given researcher, their LN is easily discoverable through such tools as Microsoft Academic Research⁶ and DBLP⁷. We show that the most productive researchers in IS have, on average, lower LNs than the general population of researchers. In other words, top researchers are likely to be close to the center of C_{IS} . This finding also holds for other researchers that are highly central in the network, such as Izak Benbasat and Jay Nunamaker; the Benbasat number and Nunamaker number would both result in the same normal distribution that we observed with Lyytinen, and these numbers would also preserve the negative correlation with productivity. We should also note that the network's center likely changes over time, which is why we display other highly connected researchers. It is heartening to note that the field's most connected researchers have a highly diverse set of interests in research topics and methods; because such researchers are brokers of knowledge transfer, the field as a whole benefits because the transferred knowledge is diverse and can enrich the field. For example, Lyytinen studies topics such as systems design, IS innovation, ubiquitous computing⁸, Nunamaker studies topics such as collaboration technology and detection deception⁹, and Argarwal studies topics such as online social networks, technology adoption and diffusion, and the impact of technology on cost reduction in healthcare¹⁰. The productivity and standing of a researcher's advisor and coauthors are likely to influence that researcher's impact on IS research (Lowry et al., 2013). Thus, one may use the Lyytinen number to assess how influential the collaborators and advisors are: a lower average LN for the author's advisor and coauthors suggests that a researcher possesses a high potential to impact the field. In this regard, the LN may prove useful for PhD students who seek academic advisors and for hiring and promotion committees as a piece of additional information when making their assessments. We do not intend the LN to replace other centrality measures, but one can use it as a piece of complementary information. A hiring committee, for example, would consider not only a researcher's LN but also that of the researcher's advisors and coauthors, which would offer a better sense of the researcher's connectedness.

Last, the eight journals in the basket of eight show a wide variety in network configuration. Compared to the other basket journals, we found that *MISQ* (52%), *ISR* (56%), and *JMIS* (47%) had relatively large maximally connected components; hence, they each had stable cores. Many sources, such as the Financial Times and UT-Dallas, view these three journals as the top IS journals (Lowry et al., 2013). Stable cores accumulate social capital for the journals. The larger the network, the more value it has (Metcalf, 1995); therefore, journals with large maximally connected components are likely to be more valuable. An alternative way of understanding the variations in connectedness among the journals is that preferential attachment is more pronounced in *MISQ*, *ISR*, and *JMIS*, which suggests that new researchers can best enter the network through collaborating with researchers with a publishing history in these journals. Second, we found higher than average collaboration levels in North American-based journals, which suggests that collaboration cultures vary across geographical regions: North American-based journals have higher levels of collaboration and larger maximally connected components. Journals with lower levels of collaboration might want to foster higher levels of collaboration in order to increase their social capitals. An example of such a venture is *ISJ*'s requirement that each research team who

⁶ <https://academic.microsoft.com>.

⁷ <http://dblp.uni-trier.de>.

⁸ <https://weatherhead.case.edu/faculty/kalle-lyytinen>.

⁹ <http://borders.arizona.edu/cms/content/jay-nunamaker>

¹⁰ <http://scholar.rhsmith.umd.edu/ragarwal/home?destination=home>

submits to its special issue on ICT4D needs to have at least one author from a developing country¹¹. As more IS journals accumulate social capital, so does the IS research field as a whole.

Table 12. Summary of Contributions

Research question	Contributions
How do author characteristics determine tie formation?	<ul style="list-style-type: none"> • We found homophily in the IS network according to gender and geography. However, we observed gender homophily only since 2000. • The proportion of female researchers has also increased. • Junior researchers prefer to collaborate with senior researchers (i.e., they collaborate on the basis of heterophily according to experience).
What are the characteristics of the IS collaboration network?	<ul style="list-style-type: none"> • Kalle Lyytinen emerged as the most central member of the network. Researchers with low Lyytinen numbers tend to be more productive on average. • The network has continued to gain connectedness since Xu et al.'s (2014) study. It is still characterized by preferential attachment: new nodes enter the network by collaborating with popular nodes. On average, it takes seven steps for one researcher to find another in the network.
How do the subnetworks corresponding to the different journals differ if at all?	<ul style="list-style-type: none"> • Three journals (<i>MISQ</i>, <i>ISR</i>, and <i>JMIS</i>) have large, dense cores. • Journals published in North America tend to be more connected than journals published in Europe. • Journals published in North America tend to have larger networks than journals published in Europe.

7 Limitations and Future Research

One must consider our results in the context of several limitations. First, the network contained 10,000 edges; we ran our analysis on 110 of these edges because manually collecting researchers' demographic data takes considerable effort and time. This sample of 110 edges was biased toward high-degree (popular) nodes. We could increase our confidence in the results with a much larger sample size. Second, we employed a RES approach to sampling the C_{IS} network. RES is better than random node sampling because it produces a subnetwork that resembles the underlying network better than the latter method; however, other more complex network sampling algorithms improve on RES. Using a sampling approach that preserves network structure, such as the hybrid random node-edge sampling method (Leskovec & Faloutsos, 2006), should reveal useful information on how structural aspects of the network such as network position influence new tie formation. Third, we define collaboration as coauthorship in the basket of eight journals. However, some instances of collaboration do not result in publication in the basket of eight or at all. Comparing such collaboration to collaboration trends in the basket of eight may further isolate the unique aspects of collaboration in our flagship journals. Fourth, and related to the third point, the sample of authors in the basket of eight might heavily favor those with prior publishing history in the journals (e.g., journal editors and their coauthors); hence, the results might display collaboration preferences of such individual groups rather than of the IS field as a whole. Fifth, we mainly examine the effects of node attributes in determining the probability that a tie exists between any two nodes. Incorporating edge attributes such as edge weight, method of research, and year of collaboration could also add extra insight into how authors choose to collaborate in the IS network. Finally, other factors such as researcher country of origin and race, complementary skill sets, similar interests, and colocation in socioeconomic hubs may also affect collaboration in the research network. Future work could add these variables to the model for validation.

8 Conclusion

Our contributions focus on the antecedents of collaboration in the IS field. Consistent with previous findings from Gallivan and Ahuja (2015), we found that gender homophily shapes coauthorship in the IS field. However, we also examined whether this homophily has held over time and found that it only became observable in the decade beginning in 2000. Moreover, we also found that female researchers select their collaborators heterophilously. Junior researchers also collaborate heterophilously with senior researchers. These findings underscore the complex nature of collaboration in the IS field because

¹¹ [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1365-2575/homepage/special_issues.htm](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1365-2575/homepage/special_issues.htm)

different groups have different opportunities for collaborating homophilously regardless of their preferences.

In summary, we examine the historical C_{IS} network by modeling it as a graph with vertices to represent authors and edges to represent coauthorship ties. We found that C_{IS} was scale free and has grown in connectedness since 2012. We created a model of collaboration and found preliminary evidence that collaboration in the IS field demonstrates gender and geographical homophily; conversely, we found evidence of field tenure heterophily. However, one must consider the results in the context of the study's limitations—particularly the fact that we conducted the analyses on only 110 edges out of a possible 10,000 edges; as we collect more data, we will gain more confidence in our findings. Lastly, Kalle Lyytinen emerged as the center of C_{IS} . We define the Lyytinen number and suggest ways that one could use it to evaluate a researcher's potential and/or past impact on the field.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions On Automatic Control*, 19(6), 716-723.
- Barabási, A.-L., & Bonabeau, E. (2003). *Scale-free networks*. American, 288(5), 50-59.
- Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, 29(4), 555–564.
- Burt, R. S. (2002). The social capital of structural holes. In M. F. Guillén (Ed.), *The new economic sociology: Developments in an emerging field* (pp. 148-190). New York: Russell Sage Foundation.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165-1188.
- Coder, L., Rosenbloom, J. L., Ash, R. A., & Dupont, B. R. (2009). Economic and business dimensions: Increasing gender diversity in the IT work force. *Communications of the ACM*, 52(5), 25-27.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5), 1-9.
- Cuellar, M. J., Vidgen, R., Takeda, H., & Truex, D. (2016). Ideational influence, connectedness, and venue representation: Making an assessment of scholarly capital. *Journal of the Association for Information Systems*, 17(1), 1-28.
- Currarini, S., Jackson, M. O., & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4), 1003-1045.
- De Castro, R., & Grossman, J. W. (1999). Famous trails to Paul Erdős. *The Mathematical Intelligencer*, 21(3), 51-53.
- Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge, UK: Cambridge University Press.
- Egorov, G., Polborn, M., & Welcome, C. A. V. (2010). *An informational theory of homophily*.
- Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization Science*, 22(5), 1224-1239.
- Feld, S. L. (1982). Social structural determinants of similarity among associates. *American Sociological Review*, 47(6), 797-801.
- Gallivan, M., & Ahuja, M. (2015). Co-authorship, homophily, and scholarly influence in information systems research. *Journal of the Association for Information Systems*, 16(12), 980-1015.
- Goh, J.-M., Gao, G., & Agarwal, R. (2016). The creation of social value: Can an online health community reduce rural-urban health disparities? *MIS Quarterly*, 40(1), 247-263.
- Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, 30(3), 611-642.
- Grossman, J. W. (2002). The evolution of the mathematical research collaboration graph. In *Proceedings of the 33rd Southeastern Conference on Combinatorics* (pp. 201-212).
- Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). Statnet: Software tools for the representation, visualization, analysis and simulation of network data. *Journal of Statistical Software*, 24(1), 1548-7660.
- Harris, J. K. (2013). *An introduction to exponential random graph modeling* (vol. 173). Thousand Oaks, CA: Sage.
- Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., & Morris, M. (2008). ERGM: A package to fit, simulate and diagnose exponential-family models for networks. *Journal of Statistical Software*, 24(3), nihpa54860.
- Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314.
- La Fond, T., & Neville, J. (2010). Randomization tests for distinguishing social influence and homophily effects. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 601–610).

- Lazarsfeld, P. F., Merton, R. K., & (1954). Friendship as a social process: A substantive and methodological analysis. In M. Berger & T. Abdel (Eds.), *Freedom and control in modern society* (pp. 18-66). New York: Van Nostrand.
- Leskovec, J., & Faloutsos, C. (2006). Sampling from large graphs. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 631-636).
- Li, S. S., & Karahanna, E. (2015). Online recommendation systems in a B2C e-commerce context: A review and future directions. *Journal of the Association for Information Systems*, 16(2), 72-107.
- Lowry, P. B., Moody, G., Gaskin, J., Galletta, D. F., Humphreys, S., Barlow, J. B., & Wilson, D. (2013). Evaluating journal quality and the association for information systems (AIS) Senior Scholars' journal basket via bibliometric measures: Do expert journal assessments add value? *MIS Quarterly*, 37(4), 993-1012.
- Mason, R. O., McKenney, J. L., & Copeland, D. G. (1997). Developing an historical tradition in MIS research. *MIS Quarterly*, 21(3), 257-278.
- McIlwee, J. S., & Robinson, J. G. (1992). *Women in engineering: Gender, power, and workplace culture*. Albany, NY: SUNY Press.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415-444.
- Metcalfe, B. (1995). Metcalfe's law: A network becomes more valuable as it reaches more users. *Infoworld*, 17(40), 53-54.
- Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987). *Rediscovering the social group: A self-categorization theory*. Cambridge, MA: Basil Blackwell.
- Venkatesh, V., Bala, H., & Sambamurthy, V. (2016). Implementation of an information and communication technology in a developing country: A multimethod longitudinal study in a bank in India. *Information Systems Research*, 27(3), 558-579.
- Walsham, G., & Sahay, S. (2006). Research on information systems in developing countries: Current landscape and future prospects. *Information Technology for Development*, 12(1), 7-24.
- Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p^* . *Psychometrika*, 61(3), 401-425.
- Watts, J. H. (2009). Allowed into a man's world' meanings of work-life balance: Perspectives of women civil engineers as "minority" workers in construction. *Gender, Work & Organization*, 16(1), 37-57.
- Wyatt, D., Choudhury, T., & Bilmes, J. A. (2008). Learning hidden curved exponential family models to infer face-to-face interaction networks from situated speech data. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence* (pp. 732-738).
- Xu, J., Chau, M., & Tan, B. C. (2014). The development of social capital in the collaboration network of information systems scholars. *Journal of the Association for Information Systems*, 15(12), 835-859.
- Yuan, Y. C., & Gay, G. (2006). Homophily of network ties and bonding and bridging social capital in computer-mediated distributed teams. *Journal of Computer-Mediated Communication*, 11(4), 1062-1084.
- Zhai, L., Li, X., Yan, X., & Fan, W. (2014). Evolutionary analysis of collaboration networks in the field of information systems. *Scientometrics*, 101(3), 1657-1677.
- Zhang, P. (2015). The IS history initiative: Looking forward by looking back. *Communications of the Association for Information Systems*, 36, 477-514.

Appendix A

Data Collection

To collect our data, we first retrieved all the tables of contents for all issues that the eight journals in the basket of eight have produced since April, 1977—the date *MISQ* began publishing. To do so, we wrote scripts that employed the UNIX scripts *wget* and *curl* to send HTTP GET requests to these journals. The journal servers honored the requests and returned files that contained the tables of contents of all journal issues published between March, 1977, and November, 2015. We placed these files into eight different directories in preparation for parsing.

To parse each HTML table of contents, we needed to use Java, regular expressions, and the Jsoup library. We parsed the files to retrieve the metadata for every published paper. We wrote a master parser that we customized for each different journal because the HTML document tree structure varied across journals. The parser outputted a map of publication titles and their respective coauthors. We wrote another Java program that created edges for each coauthor tie. Each paper with at least two authors generated $n!/2$ edges, where n is the number of authors; for example a paper with three authors A, B, and C generated $3!/2$ links – $A \rightarrow B$, $A \rightarrow C$ and $B \rightarrow C$. We did not incorporate directionality in constructing the network.

Lastly, we imported the edge lists from the last step to create 1) the individual collaboration subnetworks and 2) the combined network using the *R* programming language. We performed all subsequent calculations on the network in *R* (Ihaka & Gentleman, 1996). Figure A1 shows the process flow.

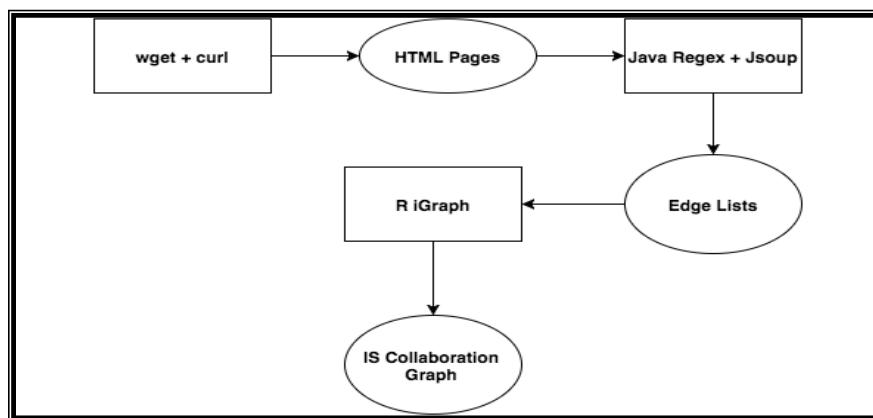


Figure A1. Method Process Flow

Technical Problems

We experienced various technical problems when collecting and parsing the data. First, two of the journal websites were not amenable to automated HTTP GET requests—possibly in fear of denial of service attacks. We had to devise workarounds in order to bypass their prevention mechanisms (in particular, through wrapping our requests with *curl* rather than *wget* for one of the websites and by reducing the average speed of our requests for the other website). Second, certain journals initially collected just the initials of the authors in lieu of their first names. We know of no automated solution to this problem; hence, we had to manually search for the first names of authors in those journals. Last, in seeking demographic information for the modeling phase (described in the ERGM subsection), certain names are very common, such as “Stephen Smith”, “Susan Brown”, and “Rui Chen”. In such cases, we used specific collaboration information to locate the correct individual for our analysis.

Appendix B: Highly Connected Researchers across Journals

Table B1. Highly Connected Researchers across Journals

Journal	Betweenness	Eigenvector
<i>EJIS</i>	Richard Baskerville	Zahir Irani
	Kalle Lyytinen	Peter Love
	Iris Junglas	Amir Sharif
	Guy Fitzgerald	P. Race
	Wynne Chin	Tony Elliman
<i>ISJ</i>	Heinz Klein	Guy Fitzgerald
	Bernd Carsten Stahl	David Avison
	Rudy Hirschheim	Rudy Hirschheim
	Frank Land	Heinz Klein
	Kalle Lyytinen	Juhani Iivari
<i>ISR</i>	Bin Gu	Detmar Straub
	Prabhudev Konana	Arun Rai
	Paul Pavlou	Edward Rigdon
	Anitesh Barua	Jan-Michael Becker
	Ritu Agarwal	Elena Karahanna
<i>JAIS</i>	Bernard Tan	Yu-wei Lin
	Varun Grover	Mark Hartswood
	Carol Saunders	Stuart Anderson
	Kalle Lyytinen	Horacio Gonzalez-Velez
	Rudy Hirschheim	Sharon Lloyd
<i>JIT</i>	Kalle Lyytinen	Hossein Zadeh
	Lars Mathiassen	Jerry Luftman
	Leslie Willcocks	Martin Santana
	Heejin Lee	Barry Derksen
	Philip Yetton	Eduardo Henrique Rigoni
<i>JSIS</i>	Leslie Willcocks	Leslie Willcocks
	Sue Newell	Mary Lacity
	Wendy Currie	Shaji Khan
	Robert Galliers	Ashok Subramanian
	Joe Peppard	Sue Newell
<i>JMIS</i>	Jay Nunamaker	Jay Nunamaker
	Olivia Liu Sheng	Robert Briggs
	Zhang Jie	Ralph Sprague
	Robert Briggs	Gert-Jan De Vreede
	Rajiv Dewan	Bruce Reinig
<i>MISQ</i>	Izak Benbasat	Izak Benbasat
	Ritu Agarwal	Paul Pavlou
	David Gefen	Alan Dennis
	Ann Majchrzak	Alok Gupta
	Wynne Chin	Rajiv Banker

About the Authors

Wallace Chipidza is a PhD candidate in the Information Systems department at Baylor University. He holds an MS in Computer Science from the University of Arizona. His work experiences include software engineer for the Apollo Group in Phoenix, Arizona, and software developer for e-Solutions in Harare, Zimbabwe. Wallace has published work in the *Journal of Computer Information Systems (JCIS)* and *Information Systems Education Journal (ISEDJ)*. His work has been included in the proceedings of the *International Conference on Information Systems*, the *Hawaii International Conference on Systems Sciences*, and the *Americas Conference on Information Systems*, among other conferences. He currently researches dynamic social networks, internet privacy, and ICT4D. He is a member of the Association on Information Systems.

Dr. John F. Tripp is Assistant Professor of Information Systems at Baylor University. Before beginning his PhD, he worked for more than 17 years in industry as a software developer, project manager, and IT Director. Dr. Tripp completed his PhD at Michigan State University in 2012. His research has appeared or is forthcoming in the *Journal of the Association for Information Systems*, *Computers and Human Behavior*, *Journal of Computer Information Systems*, *Information Systems and e-Business Management*, and the *Journal of Management Systems*. His research has appeared in the proceedings of multiple conferences including the *International Conference on Information Systems*, the *Hawaii International Conference on Systems Sciences*, the *European Conference of Information Systems*, and the *Americas Conference on Information Systems*. He lives in McGregor, Texas, with his wife and eight children.

Copyright © 2018 by the Association for Information Systems. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and full citation on the first page. Copyright for components of this work owned by others than the Association for Information Systems must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior specific permission and/or fee. Request permission to publish from: AIS Administrative Office, P.O. Box 2712 Atlanta, GA, 30301-2712 Attn: Reprints or via e-mail from publications@aisnet.org.