

**Mathematics, Statistics and Computer Science Faculty Research and Publications/College of Arts and Sciences**

***This paper is NOT THE PUBLISHED VERSION; but the author’s final, peer-reviewed manuscript.***

The published version may be accessed by following the link in the citation below.

*Journal of the American Society for Information Science and Technology*, Vol. 61, No. 4 (January, 2013): 162-172. [DOI](#). This article is © Wiley and permission has been granted for this version to appear in [e-Publications@Marquette](#). Wiley does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Wiley.

**Contents**

Abstract.....	2
Introduction .....	3
Background and Literature Review.....	4
One Institution, Two Campuses: Collaboration Trends .....	4
Individual Collaborators.....	4
Institutional Support .....	5
Research Context and Methods.....	6
Results.....	6
RQ1: Overall Patterns of Cross-Campus Collaboration.....	6
Departmental CC collaboration activity.....	7
Citation impact of SC and CC collaboration .....	9
Author distribution across campuses .....	9
RQ2: Who Is Involved in CC Collaboration? .....	9
Who are the best connected cross-campus collaborators? .....	10
RQ3: Institutional Support for CC Collaboration.....	12
Discussion.....	13
Implications.....	15

Limitations and Future Work .....	16
Acknowledgments.....	16
Footnotes.....	16
References .....	17

# Cross-campus collaboration: A scientometric and network case study of publication activity across two campuses of a single institution

**Jeremy Birnholtz**

Department of Communication, Department of Information Science, Cornell University, Ithaca, NY

**Shion Guha**

Department of Information Science, Cornell University, Ithaca, NY

**Y. Connie Yuan**

Department of Communication, Cornell University, Ithaca, NY

**Geri Gay**

Department of Communication, Cornell University, Ithaca, NY

**Caren Heller**

Department of Public Health, Weill Cornell Medical College, New York, NY

## Abstract

Team science and collaboration have become crucial to addressing key research questions confronting society. Institutions that are spread across multiple geographic locations face additional challenges. To better understand the nature of cross-campus collaboration within a single institution and the effects of institutional efforts to spark collaboration, we conducted a case study of collaboration at Cornell University using scientometric and network analyses. Results suggest that cross-campus collaboration is increasingly common, but is accounted for primarily by a relatively small number of departments and individual researchers. Specific researchers involved in many collaborative projects are identified, and their unique characteristics are described. Institutional efforts, such as seed grants and topical retreats, have some effect for researchers who are central in the collaboration network, but were less clearly effective for others.

## Introduction

Collaboration across disciplinary and geographic boundaries has proven critical to solving important problems, and there is substantial evidence of a shift from single investigators and authors to team responsibility for knowledge generation (Falk-Krzesinski, Börner, et al., 2010; Wuchty, Jones, & Uzzi, 2007). In particular, collaboration can provide expertise in multiple areas (Disis & Slattery, 2010) and the ability to pool resources to enable the massive projects often called “big science” (Galison & Hevly, 1992). The desire to collaborate is also influenced by factors such as institutional constraints (Landry & Amara, 1998), the availability of “attractive” collaborators in terms of influence or unique skills (Bozeman & Corley, 2004; Hara, Solomon, Kim, & Sonnenwald, 2003; Shrum, Genuth, & Chompalov, 2007), entrepreneurial aspirations (Oliver, 2004), attributes of the work to be performed (Birnholtz, 2007), and the need for access to special data or research equipment (Beaver, 2001; Melin, 2000).

Moreover, collaboration across geographic distance has been enhanced substantially by new communication and collaboration technologies that have spawned new styles of work, communication, and coordination (Cummings, Finholt, Foster, Kesselman, & Lawrence, 2008; Nentwich, 2003; Olson, Zimmerman, & Bos, 2008). At the same time, there is also substantial evidence that collaboration is difficult. Science continues to operate on an economy of individual reputation (Whitley, 2000)—it is individuals who are awarded tenure, prizes, and other markers of status—and collaboration can make it difficult to discern how to attribute credit or blame (Birnholtz, 2006). Some researchers have also questioned the use and feasibility of collaboration (“Who'd want to work in a team?” 2003; Kennedy, 2003). Moreover, funding agencies and institutions have struggled in their efforts to encourage effective research collaboration (e.g., Koku & Wellman, 2004). Projects spanning multiple universities have also been plagued by problems with coordination and interpersonal conflicts (e.g., Cummings & Kiesler, 2007).

There has been less study, however, of research collaborations that span multiple campuses of a single institution. One notable example is Lee, Brownstein, Mills, and Kohane's (2010) recent study suggesting that geographic proximity among collaborators can positively affect the citation impact of research. Although many of the difficulties listed above—such as distance and differences between disciplines—still apply, a single university is also uniquely situated to encourage and support collaboration via communication infrastructure, seed grants, and other financial incentives, social opportunities for like-minded researchers to connect with each other, and administrative structures that can simplify coordination across sites and units.

Although at first glance collaboration among multiple campuses of one institution may seem easier than collaborations across institutional boundaries, there are nonetheless numerous obstacles (Birnholtz et al., 2012). In the article that follows, we present a case study of cross-campus collaboration at Cornell University, using single-campus collaboration within the same institution as a comparison case. Cornell is a university where the medical campus is located in New York City, approximately 200 miles away from the rest of the university, which is in Ithaca, New York. We use authorship and citation data for Cornell faculty to compare and evaluate the cross- and single-campus collaborations.

## Background and Literature Review

### One Institution, Two Campuses: Collaboration Trends

Collaborations that span multiple campuses of a single university can bring together those with expertise around important problems and exploit the benefits of a shared institutional context (Birnholtz et al., 2012). This can be particularly important in addressing problems in the life sciences (Disis & Slattery, 2010), which often involve medical school faculty that may be located in a separate facility with its own distinct administrative structure and often on another campus. By combining the clinical experience and knowledge of medical researchers with experts in fields such as biomedical engineering, physics, and others, there are opportunities to develop novel technologies that can address significant medical problems. There are also significant opportunities for federal funding (Huerta et al., 2005) in this area, along with opportunities for novel intellectual property (Owen-Smith & Powell, 2003). Universities have recognized this, and many have sought to encourage collaboration in these areas via the construction of new facilities and development of collaborative programs (Schwenk & Green, 2006).

One problem with sparking collaboration across campuses, however, is that collaborators located several hundred miles apart may not know or be aware of each other, and may be wary of engaging with remote collaborators they do not know well (Nomura et al., 2008). Institutions can take some steps to encourage these projects, and this has been referred to as the “support” component of team science (e.g., Falk-Krzesinski et al., 2010). There are significant open questions, however, about how to effectively provide institutional support for team science. Simply putting incentives in place, for example, neither leads researchers to want to collaborate (“Who'd want to work on a team?” 2003), nor necessarily leads to effective or impactful collaboration (Cummings & Kiesler, 2007).

In determining how to better support cross-campus collaboration, it is therefore useful to examine how frequently collaboration occurs generally, and how collaboration behavior is distributed across departments on both campuses of the university. It is further useful to examine how cross-campus collaboration activity compares with single-campus activity. It could be the case, for example, that cross-campus collaboration opens up new possibilities for collaboration to occur where it did not before, or that those who already collaborate frequently will do so more over these boundaries. These analyses will allow for better understanding of where to focus efforts to encourage collaboration.

A desire to understand the frequency, composition, and impact of cross-campus collaborations motivated our first research question:

**RQ1:** How frequently and in what departments does single- and cross-campus collaboration occur at Cornell? What is the citation impact of the work from both types of collaboration?

### Individual Collaborators

Decisions about academic collaboration rest fundamentally at the level of the individual researcher (Bozeman & Corley, 2004). Individual researchers, however, work within social and professional networks. It has been shown repeatedly that some individuals in these networks tend to be more central than others, acting as hubs for information (Velden, Haque, & Lagoze, 2010) and as brokers helping people with similar interests meet each other. Such links can be particularly important in research networks and for establishing collaborations, and in the sharing of data, information, and other resources. Many have examined networks of coauthors, ranging from Crane's (1972) early observation

that research tends to occur in “invisible colleges,” to more recent examinations that identify particularly central or influential researchers in specific fields such as physics (Newman, 2001) and human-computer interaction (Horn, Finholt, Birnholtz, & Motwani, 2004).

This raises another key challenge in the initiation of research teams: the establishment of social ties with potential collaborators. Beyond simply being aware of potential collaborators and their expertise, gaining access to others' knowledge and expertise (and ultimately working together in a team) depends on social ties. Although developing a sense of “who knows what” is important for locating needed experts, this is only a necessary, but not a sufficient condition for successful collaboration (Yuan, 2009; Yuan, Fulk, Monge, & Contractor, 2010). Yuan and her colleagues found that the strength of communication ties mediated the relationship between awareness of who is an expert in a field and access to expert information. This means that even when team members were aware of who had the needed expertise, without the support of social ties, simply knowing “who knows what” did not always result in actual access to needed expertise.

By looking at networks of coauthors on collaborative publications and identifying individuals central to these networks, we can identify specific individual researchers that appear to be playing key roles in facilitating collaboration between individuals, such as brokering social relationships (Yan & Ding, 2009) or engaging in large numbers of collaborative projects. We can then learn more about these individuals to see what factors affect their collaboration behavior, as well as the quantity and impact of their work. These possibilities motivated our second research question:

**RQ2:** What factors predict centrality in the network of cross-campus authors? Which authors are central/highly productive in cross campus collaboration and what are their unique attributes?

## Institutional Support

Finally, we wondered about the effects of institutional efforts to encourage collaboration more generally, and cross-campus collaboration in particular. As discussed earlier, we wonder about what practices the administration can adopt to support effective interdisciplinary collaboration that cannot be done when multiple institutions are involved. In other words, what can be done to turn the potential benefits of belonging to one institution into a real advantage?

As noted above, it is often the case that potential collaborators in separate locations may be unaware of each other's skills and interests (Nomura et al., 2008). Some institutions have taken explicit steps to make these researchers aware of each other's interests, skills, and expertise. This can be accomplished via expertise referral systems (e.g., Contractor, Zink, & Chan, 1998; Shami, Ehrlich, & Millen, 2008), retreats focused on particular topics (Nomura et al., 2008), workshops, institutes, and institutional structures that encourage interdisciplinary collaboration and by making seed funding available for pilot projects.

To assist with this, Cornell established an office of intercampus initiatives in Ithaca in 2005 and in New York City at the medical school in 2006. The office is charged with helping to foster cross-campus collaboration. The office has taken several steps including seed-funding grants awarded competitively to researchers and retreats organized by the administration and lead researchers, both generally and in key targeted areas, to bring researchers from the two campuses together around a series of presentations and discussions. We wondered about the relationships between these efforts, the amount of cross-campus collaboration activity, and the centrality of individuals involved. Thus:

**RQ3:** What are the relationships between institutional support initiatives (e.g., funding, retreats), cross campus collaboration activities, and a faculty's involvement in across-campus collaboration?

## Research Context and Methods

To analyze and describe single- and cross-campus collaboration, we use publication and coauthorship data because they provide a uniform measurement and are also a key dimension used in evaluating faculty performance. Authorship data were first collected from a database shared with us by Cornell University's VIVO project team (VIVO Collaboration, 2010). VIVO is a national project funded by the National Institutes of Health (NIH) that aims to connect researchers from multiple disciplines (VIVO Collaboration, 2010). Institutional affiliation (i.e., job title, graduate field membership, etc.) data were extracted for the 3,543 full-time tenured and tenure-track faculty employed by Cornell University in October 2010.

For each individual included in the VIVO database, publication and citation data were then extracted from ISI Web of Knowledge (ISI Web of Knowledge, 2011) using software developed in Python by the project team. Data were extracted for 17,872 articles published between 1999 and 2009 with at least two Cornell-affiliated authors. This 10-year window was chosen to capture activity both before and after the creation of the Office of Inter-Campus Initiatives, and 2009 was selected as an endpoint to allow time for citations to accrue, so that citation impact could reasonably be examined.

One common problem in extracting publication data are that many authors have similar names or initials, resulting in ambiguity about whom a particular article should be attributed to (Elliott, 2010). Our automated extraction from the ISI database resulted in 686 articles for which author information was ambiguous. To resolve these instances, each author's name was manually compared to the author list extracted from VIVO, and publications were verified by checking the author's individual or departmental website. No ambiguous cases remained after this process.

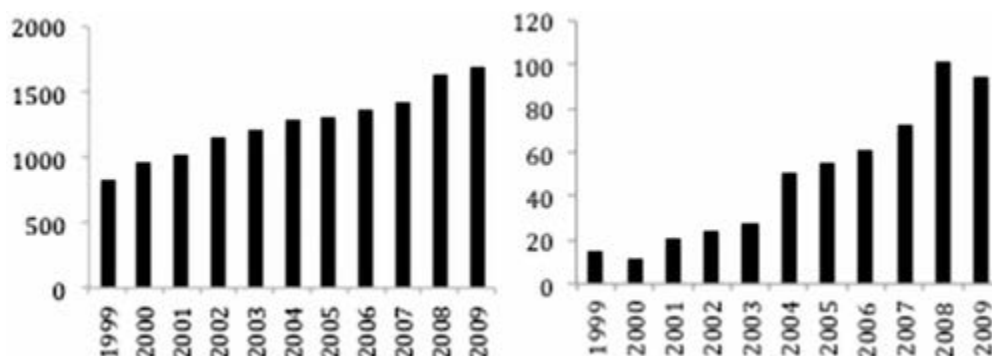
Within the full set of extracted article data, we identified 543 that had authors from both of Cornell's campuses, which we considered our CC (cross-campus) set of publications. The authors on these papers form what we refer to as the cross-campus collaboration network. An individual's home campus was determined by their primary departmental affiliation, in consultation with the Offices of Inter-Campus Initiatives (with which one of the authors of this paper is affiliated). The remaining 17,329 articles had authors from one campus, and were considered our SC (single campus) set. We identified 549 individuals with at least one CC article in the data set. Institutional data regarding seed grant funding and attendance at retreats were then systematically collected from archival records kept since 2006.

## Results

### RQ1: Overall Patterns of Cross-Campus Collaboration

We were first interested in how often CC and SC collaboration were occurring, and whether there has been growth over the period being studied. The raw quantities of CC vs. SC papers presented above make it clear that CC collaboration is less prevalent than SC, but we wondered if CC had grown at a faster rate given recent interests in fostering team science, and institutional efforts within Cornell to span the two campuses.

The numbers of SC and CC papers published each year are presented in Figure 1. As expected, the total number of SC publications is far greater over the entire time period. When the trends in growth rates are compared, however, this becomes more interesting. There appears to be a steady growth in SC collaboration (Figure 1, left). Turning to CC collaborations (Figure 1, right), some growth also occurred but the trend appears to be less steady. There are clear bursts of activity around 2004 and 2008, and a slight drop in 2009. The increase after 2005 likely reflects increased institutional interest in fostering CC projects. It is not clear whether the drop in 2009 represents the start of a downward trend or possibly reflects the limited availability of seed funding in 2008 and 2009, or the start of a plateau or stabilization. Regardless, however, there has been a clear surge in CC collaboration over the period being examined.



**Figure 1** Histograms for number of single-campus articles (1999–2009; left). Histogram for number of cross-campus articles (1999–2009; right).

### Departmental CC collaboration activity

Having looked at the overall incidence and growth of CC collaboration, we wanted to understand the distribution of this activity across departments on both campuses. For each department with at least one CC publication ( $N = 88$  of 170 total departments), we examined: (a) the total number of CC and SC articles published, (b) the total number of authors on each article, and (c) the number of citations per article.

Looking first at the total number of CC articles published by faculty in each department, the mean total number of CC articles per department is 3.65 ( $SD = 3.02$ , median = 2.5, mode = 1), which is much lower than the number of SC articles per department ( $M = 81.13$ ,  $SD = 104.95$ , median = 30.5, modes = 26,50). Thus, CC articles account for approximately 5% of total collaborative publication activity per department, with the modal frequency even lower.

Looking across departments, however, we found that a small number of departments account for a relatively large number of CC articles across the university. We therefore decided to look more closely at the 20 departments producing the greatest numbers of CC publications. These 20 departments, rank ordered in Table 1, constitute 11% of all departments, but account for approximately one third (30.75%) of all CC articles published ( $N = 543$ ) and approximately one third (30.61%) of all authors who have published CC articles ( $N = 549$ ). As expected, there were more CC articles per department in these 20 departments ( $M = 8.35$ ,  $SD = 2.53$ , median = 8, mode = 8) over our 10-year window, than those in the remaining 68 departments ( $M = 2.28$ ,  $SD = 1.25$ , median = 2, mode = 1),  $t(87) = 10.69$ ,  $p < 0.05$ .

Interestingly, the same was true in this same Top 20 group for SC articles ( $M_{Top\ 20} = 214.71$ ,  $SD = 111.96$ , median = 218.50, mode = 257;  $M_{Others} = 41.84$ ,  $SD = 62.1$ , median = 11.50, mode = 8),  $t(87) = 6.91$ ,  $p < 0.05$ .

**Table 1.** Articles and authors for the 20 departments with the most CC articles

Department	Campus	Total number of faculty	CC articles produced		Number of faculty with at least one CC article	
			Rank	Quantity	Rank	Quantity
Psychiatry	NYC	37	1	17	5	8
Medicine	NYC	517	3	10	1	17
Microbiology & Immunology	NYC	44	2	11	4	9
Neurology	NYC	17	3	10	4	9
Surgery	NYC	26	4	9	2	11
Radiology	NYC	33	4	9	2	11
Molecular Biology	NYC	28	4	9	4	9
Physics	ITH	43	5	8	10	8
Electrical & Computer Engineering	ITH	33	5	8	4	9
Biomedical Engineering	ITH	11	5	8	3	10
Physiology & Biophysics	NYC	110	5	8	5	8
Population Medicine	ITH	13	5	8	5	8
Mechanical Engineering	ITH	41	5	8	5	8
Pharmacology	NYC	18	6	7	7	6
Computer Science	ITH	29	6	7	6	7
Economics	ITH	26	6	7	8	5
Biochemistry	ITH	18	7	6	6	7
Biomedical Sciences	ITH	20	7	6	7	6
Cell & Development Biology	NYC	16	7	6	7	6
Reproductive Medicine & Infertility	NYC	14	8	5	6	7

We considered the possibility that department-level productivity was purely a function of department size. Only seven of the top 20 CC departments, however, are among the top 20 departments by size (in terms of the number of faculty). This gives us some indication that department size is not critically confounding these results. We also noted that the top 20 departments are nearly evenly distributed across the two campuses, with nine located on the Ithaca campus, and the remaining 11 in New York City.

All of this suggests that CC collaboration tends to occur in departments that also engage in SC collaboration, as opposed to CC collaborations affording collaboration opportunities to researchers who would not otherwise work with others.



## Citation impact of SC and CC collaboration

Next, we examined the citation impact of CC collaborations as compared with SC, by comparing the number of citations per article for both SC and CC publications, using ISI citation data. SC and CC citation counts were compared within departments to avoid possible confounding by documented differences between disciplines in publication and citation patterns. On average, departments producing one or more total CC articles had 15.16 ( $SD = 18.14$ , median = 9.90) citations per CC article, and 8.06 ( $SD = 8.03$ , median = 5.85) citations per SC article, which a Welch's paired-sample  $t$  test (assuming unequal variances and sample sizes) shows is a statistically significant difference,  $t(87) = 3.31$ ,  $p < 0.01$ . Thus, citation impact for CC articles in our data set is greater than that for SC articles. This suggests there is potential value in CC collaboration in that the research may be exposed to broader audiences or have more intellectual value.

## Author distribution across campuses

We also looked at the proportion of authors on each CC article from both campuses. We found that, of 549 total authors on CC articles, 66.6% are from the New York City campus, and 33.3% are from Ithaca. On average, there were 2.15 authors from New York City on each CC article, and 1.24 from the Ithaca campus. This imbalance, although not tested statistically, was striking in that the majority of faculty is affiliated primarily with the Ithaca campus. It suggests that at least some CC collaborations may involve existing groups on the New York City campus reaching out to single collaborators on the Ithaca campus, or lone researchers in Ithaca seeking groups of collaborators on the medical campus. Both of these are in opposition to a scenario in which similarly sized groups join together, in which case we might expect the numbers to be more even on the two campuses. In some ways, however, this is not surprising as researchers on the Ithaca campus have a larger number of potential collaborators readily available to them. More broadly, this means that efforts to encourage collaboration should focus both on group–group collaboration opportunities as well as group–individual opportunities.

## RQ2: Who Is Involved in CC Collaboration?

From the department-level data discussed above, it became clear in our analyses that—even in the most active departments—a relatively small fraction of the total faculty were responsible for much of the observed CC activity. To better understand these individuals and patterns of CC collaboration, we turned to the properties of the CC coauthorship network. As discussed earlier, the CC coauthorship network is the set of faculty who were authors on at least one CC article in our data set. A link in the CC network is present in the network when two people coauthored a publication. For each of these individuals, we used UCInet 6.0 (Borgatti, Everett, & Freeman, 2002) to extract three normalized centrality metrics: degree, betweenness, and closeness. We use these measures to understand which faculty is more influential in the coauthorship network. Here, the word “influential” refers to any of the relative interpretations of the centrality metrics of a network. Degree centrality measures the number of coauthors that a person has across all of his/her publications. Betweenness centrality measures the extent to which an author can function as a broker of coauthorship to get other authors in the network connected (Bonacich, 1987; Freeman, 1977). Unlike degree centrality, which focuses on direct coauthorship only, closeness centrality measures whether an author is well connected to the other authors in the collaboration network through both direct and indirect collaboration ties.

We used regression models to identify factors predicting network influence. We examined the individual's primary campus location (NYC or not), academic rank (assistant, associate, or full professor),

position classification (research, clinical, or neither), and the number of Cornell graduate fields<sup>2</sup> to which an individual belongs. To test these relationships, we used a generalized gamma regression model<sup>3</sup> run in a backwards-stepwise mode. Results from these models are presented in Table 2.

**Table 2.** Generalized gamma regression model of the relationship between centrality and author trait variables

Variable	Model A parameter estimates (betweenness centrality as DV)	Model B parameter estimates (closeness centrality as DV)
Fields	0.55a	0.84a
Location	0.66	1.04
Faculty Rank	1.07	0.94
Position Type	1.01	0.93

Note.  $df = 547$ .

a  $p < 0.05$ , two-tailed.

As the table shows, only the number of graduate field memberships is a significant predictor of the normalized betweenness centrality ( $B = 0.55$ ,  $t = 1.76$ ,  $p < 0.05$ ) and closeness centrality ( $B = 0.84$ ,  $t = 1.95$ ,  $p < 0.05$ ). Thus, an individual with a higher number of field affiliations is more likely to function as a “broker” in the coauthorship network, and does not require as many “hops,” on average, to reach another author in the network. We must note here that the distribution of “fields” is highly skewed (Median = 1, Mode = 1), ranging from 1 to 4. Thus, an author with two field memberships has more than most authors. Our regression model, which accounts for the skewed distribution, suggests that on an average CC authors with more than one field membership have higher normalized betweenness and closeness centrality values.

A similar model with the normalized degree centrality metric failed to converge because no relationships could be found between this measure and the independent factors.

### Who are the best connected cross-campus collaborators?

We next looked more carefully at the 10 individuals with the highest centrality scores in the network (see Table 3).

**Table 3.** Top 10 ranked list of authors by centrality metrics

Author	Department	Degree	Betweenness	Closeness	Location
Wang	Biomedical Engineering	1	1	1	ITH
Prince	Radiology	2	48	16	NYC
Lee	Surgery	3	6	3	NYC
Felsen	Urology	4	57	82	NYC
Kim	Radiology	5	3	2	NYC
Poppas	Urology	6	58	83	NYC
Jain	Pediatrics	7	24	12	NYC
Chen	Psychiatry	8	28	5	NYC
Kent	Surgery	9	98	61	NYC
Nguyen	Radiology	10	146	40	NYC

As shown in Table 3, Wang had the top centrality scores across all three measures. Prince, Felsen, and Poppas, in contrast, scored high on degree centrality, but much lower on betweenness and closeness centrality. Through conversations with faculty and administrators, we learned that Wang and Prince work in the same lab space, and have collaborated with each other frequently—hence the high degree centralities for both of them. Prince, however, has collaborated almost exclusively with Wang, whereas Wang also collaborated with other Ithaca researchers. This explains the difference between them on the other centrality measures.

We also learned that Felsen and Poppas formerly collaborated with each other almost exclusively, which again explains their lower scores on the centrality measures that gauge a person's capacity for brokering collaboration or reach others for collaboration in the network. Similar to Wang, Lee collaborated with different people from different departments on CC projects, and hence received high scores along all dimensions too.

As can be seen from the additional descriptive statistics presented in Table 4, the most central faculty in the CC coauthorship network are similar to each other, but different in important ways from most faculty, including those with CC publications. On average, the 10 most central CC authors, for example, have more than twice as many CC ( $M = 54.5$ ,  $SD = 27.53$ ) publications as they do SC ( $M = 21.2$ ,  $SD = 13.72$ ),  $t(9) = -3.99$ ,  $p < 0.01$ . By contrast, most faculty have more SC publications than CC; the mean number of CC publications for individuals who have published at least one CC publication is 8.89 ( $SD = 14.91$ ), and the mean number of SC publications in this same group is 26.87 ( $SD = 27.66$ ). As was true for departments producing CC work, the most central CC authors also had their CC publications receive more citations ( $M = 19.82$ ,  $SD = 9.65$ ) than their SC publications ( $M = 12.83$ ,  $SD = 6.59$ ),  $t(9) = -4.89$ ,  $p < 0.01$ .

**Table 4.** Descriptive statistics for the top 10 most central authors

Author	Total SC articles	Total CC articles	Citations / SC article	Citations / CC article	Funding received	Retreat attended	Fields
Wang	54	67	23.41	39.75	Yes	Yes	2
Prince	34	66	17.58	18.64	Yes	Yes	1
Lee	15	53	18.47	23.67	Yes	Yes	1
Felsen	14	86	8.28	11.94	Yes	Yes	1
Kim	11	14	3.76	7.82	Yes	Yes	2
Poppas	15	85	17.93	25.61	Yes	Yes	2
Jain	22	67	9.56	15.83	No	No	2
Chen	8	38	4.75	8.84	No	No	2
Kent	24	64	15.33	26.74	Yes	Yes	1
Nguyen	15	5	9.21	19.33	Yes	Yes	1

Another finding from Table 4 that surprised us is that, unlike the others, two of the top collaborators: Chen and Jain, did not attend retreats or receive seed funding. We learned, however, that both may have been involved with cross-campus education initiatives that we did not explicitly measure but could have helped build cross-campus ties. These cases, combined with statistics on field membership, point to the importance of graduate student education in building interdisciplinary relationships across the campuses.

These top 10 authors account for 255 CC articles, which comprise 47% of the total set of CC articles. This strongly suggests that these were the major “super-collaborators” at Cornell University during the period studied.

### RQ3: Institutional Support for CC Collaboration

Finally, Cornell has taken several steps to encourage CC collaboration, including (a) seed-funding grants awarded competitively to researchers and (b) retreats organized by the administration and lead researchers, both in general and in key targeted areas to bring researchers from the two campuses together around a series of presentations and discussions. We wondered about the broader relationship between these efforts and CC publication productivity and impact. We first present descriptive statistics for the average citations per cross campus article for both funded authors and retreat attendees.

Several interesting findings emerge. First, only 21 authors (28.37% of the total funded authors) who received seed funding for inter-campus collaboration published at least one CC article between 1999–2009, whereas 74 (13.47% of all CC authors) authors who received seed funding published at least one SC article during the same time period. On average, those who received seed funding published 26.86 SC and 8.89 CC articles during this time period, as compared with the overall average of 27.91 SC and 7.92 CC articles during this time period. Similarly, those who attended retreats published 29.93 SC and 5.86 CC articles during this time period as compared with the overall average of 27.91 SC and 7.92 CC articles during this time period.

Turning next to the impact of these articles, we conducted two paired-sample *t* tests (two-tailed and one-tailed). As Table 5 shows, mean citations per CC article ( $M = 26.41$ ,  $SD = 18.04$ ) are significantly greater than mean citations per SC article ( $M = 13.97$ ,  $SD = 10.13$ ) for funded authors,  $t(20) = 2.01$ ,  $p = 0.03$ . This suggests that, for those who received seed funding and those who actually published a CC article, their CC collaborations yielded higher impact work than their single-campus articles.

**Table 5.** Citations for single- and cross-campus articles for those who attended retreats and receive seed funding

	Single Campus Articles		Cross Campus Articles	
	Mean	SD	Mean	SD
Citations per Article				
Funded Faculty	13.97 <sup>a</sup>	10.13	26.41 <sup>a</sup>	18.04
Non-Funded Faculty	14.15 <sup>a</sup>	11.99	16.51 <sup>a</sup>	17.80
Citations per Article				
Retreat Attendees	12.40 <sup>a</sup>	8.75	17.27 <sup>a</sup>	21.66
Non-Attendees	15.77 <sup>a</sup>	12.31	14.59 <sup>a</sup>	19.80

*Note.* Total N varies due to variation in the number of authors in the two categories who had published both SC and CC work. N indicated by superscripts as follows: <sup>1</sup>N = 74, <sup>2</sup>N = 21, <sup>3</sup>N = 475 <sup>4</sup>N = 58, <sup>5</sup>N = 40, <sup>6</sup>N = 491.

We then examined those who attended retreats. Of all retreat attendees, 40 (68.96%) have published CC articles. Note that 18 individuals both received seed funding and attended retreats. We then conducted similar paired-sample *t* tests to compare the average citations/article for both SC and CC articles for retreat attendees, which are seen in the bottom two rows of Table 5. Results do not show a significant difference between the average citations/article for both SC and CC articles for retreat attendees,  $t(39) = 1.26$ ,  $p > 0.10$ . Thus, attending retreats did not seem to correlate with higher impact

CC or SC publications. We also conducted Welch's *t* tests for unequal variances and sample sizes to compare collaboration impact across groups (funded vs. nonfunded and retreat vs. nonretreat). However, no statistically significant difference in any of the tests was found.

Looking back at Table 4, we see further that 80% of the 10 most central CC authors received seed funding for CC projects, and that 80% of them attended at least one focused retreat aimed at encouraging CC projects. This is substantially different from the broader population of CC authors, in which only 10.5% attended at least one retreat, and 13.4% received seed grant funding.

We also considered the sequencing of retreats and seed grants received relative to CC article production. Unfortunately, detailed information about the timing and funding source for each article is not available to us, so we cannot make statistically rigorous causal statements about this. However, we do notice that the majority of the seed grants and retreats were organized in the time period 2005–2007. There is a general rise in CC article activity through and after this time period in our data. Specifically, more CC articles were produced in 2008 and 2009 by authors who either received seed grants or attended retreats than in earlier time periods. The broader effects of these efforts are discussed in the next section.

## Discussion

We began this case study with questions about the patterns and nature of CC collaboration at Cornell University. RQ1 examined the overall patterns of CC collaboration at Cornell. The data suggest that CC collaborations represent a relatively small fraction of overall research collaboration, but that they occur regularly and have become more common over the past several years. CC collaborations are larger in scale than SC collaborations, on average, likely because they tend to be interdisciplinary projects. Interestingly, CC collaborations also tend to include more individuals from the medical (NYC) campus than the Ithaca campus, which is contrary to expectations given the overall faculty size on both campuses. This suggests that those seeking to encourage collaboration should look at opportunities both for bridging existing research groups, as well as for creating new groups by adding remote individuals to existing groups. In addressing our earlier question about whether CC collaboration opens up new avenues for collaboration to those who otherwise would not work together, our results—particularly for the top CC collaborators, who have a large number of CC papers (and fewer SC papers)—suggest that CC collaboration provides valuable new opportunities for collaboration.

Not surprisingly, the distribution of CC collaborations among departments is uneven. Only about half of the 170 departments had at least one CC publication, and there were only 3.25 CC publications per department, on average, for the 10-year window we examined. When we examined the top 20 most productive CC departments, however, we found that they had nearly four times more CC publications on average (8.35 vs. 2.28), and these 20 departments accounted for more than 30% of the CC publication activity overall. The top 20 departments also had nearly twice the fraction of faculty involved in CC work (23.5% vs. 12.35%). Interestingly, the citation rate for CC articles was also found to be higher than for SC, suggesting the utility of inter-disciplinary research from an impact standpoint. It was also the case that CC publications from the top 20 CC departments had a higher citation rate than those from the remaining departments. Thus, it seems clear that some departments are particularly well suited to CC collaboration. They publish more CC papers, have a larger proportion of their faculty involved in CC collaborations, and publish CC publications that are cited more. These departments are clear targets for

encouraging future collaboration, as they are home to experienced collaborators and topically relevant work. Given the list of top CC departments and our own knowledge of the medical research context, part of the reason for this is topical. There are clear points of overlap in the interests of researchers in these departments, and many of these have developed into fruitful collaborations. This raises two key questions for future investigation.

The first is a question of CC saturation. Is there, in other words, a point at which the bulk of the fruitful CC collaboration opportunities in a given department have been tapped? Given the difficulties of initiating and sustaining CC collaboration (e.g., Birnholtz et al., 2012; Cummings & Kiesler, 2007) and the benefits of having geographically proximate collaborators (Lee et al., 2010), it is important to consider the balance of when it is useful to encourage more CC collaboration in a particular area, versus spreading to other areas.

This raises the second question of how to identify other areas ripe for future CC collaboration opportunities. Our results show that motivated individuals—even without institutional support—can be effective in starting new projects and working with collaborators on another campus. These are signs of possible additional opportunities that are worthy of future institutional support. Our results also show small amounts of CC activity in a wide range of departments including, for example, the social and behavioral sciences. It is difficult to tell based on these data whether these are isolated incidents or hold the potential for future collaboration. Additional qualitative analysis would be useful in understanding how to assess potential in untapped areas.

As discussed previously, it was clear from our data that a relatively small fraction of individuals in any given department were involved in CC research, so we explored factors that predict the network centrality and productivity of individual researchers. Interestingly, the number of graduate field memberships for any given faculty member—which is a unique institutional property of Cornell—was the only significant predictor of centrality in our models. This suggests that although it is possible that highly central individuals would be involved in CC collaborations whether or not the field membership structure existed, the results from our statistic test showed that the structure of Cornell's graduate school, which is intended to facilitate interdisciplinary collaboration, is having the desired effect.

We then examined the characteristics of the 10 most influential individuals to better understand what distinguishes them from the others. Most notably, these 10 individuals account for nearly half (47%) of all CC publications in the time window we examined. They were different from other faculty in that they were far more likely to have attended a topical retreat, to have received seed funding for a CC project, and to have more CC than SC publications. Interestingly, however, their rates of citation did not differ significantly from the broader faculty.

We also wondered, though, if the seed funding and retreat programs at the institutional level were also benefiting other faculty, and leading to CC research. Results here were less clear. Only a relatively small fraction (13.47%) of those who have received seed funding have published CC articles, though it is possible that some of these projects have not yet yielded publishable results. Thus, it is not immediately clear that these institutional programs are having a broad impact. Meanwhile, a higher fraction (68.96%) of retreat attendees have published at least one CC article, particularly those who are highly central individuals, suggesting that retreats can be effective institutional interventions to encourage CC collaboration.

## Implications

These results have several implications for those interested in understanding or encouraging research collaboration within a single institution. First, it seems clear that there is substantial interest in CC collaboration, as it is occurring with greater frequency, and that institutional efforts to facilitate CC collaboration seem to be having some effect. One thing to keep in mind, however, is that the distribution of these efforts is not at all even across departments, or even across faculty within departments. There is clear evidence that relatively small numbers of departments, and faculty within those departments, are accounting for a substantial fraction of the CC activity. Institutional efforts, such as retreats and seed funding, seemed to be most effective for those individuals who are highly central in the cross-campus coauthorship network.

Although funding opportunities and retreats are effective strategies to support both single and multi-institutional projects, one potential limitation of such strategies is that they operate on cycles. A more “steady” strategy found in our case study on Cornell is its assignment of field memberships that cut-across departmental lines. The finding that central CC authors, on average across the entire data set, tend to have a greater number of field memberships showed the value of administrative intervention in supporting day-to-day cross-disciplinary interactions and in conveying the institution's commitment to cross-disciplinary collaboration. At least two lines of existing management theories and research can be used to explain why multiple field memberships are valuable for fostering CC collaboration. First, research on structural hole theory (Burt, 2001) shows that those who connect disparate subgroups in a network enjoy greater information benefit and thereby achieve better performance. Second, as found in multiple studies on transactive memory systems in organizations (Lewis, Lange, & Gillis, 2005; Yuan, Carboni, & Ehrlich, 2010), accurate knowledge of expertise distribution is essential for employees to better coordinate expertise and to achieve good performance. To develop such accurate knowledge requires employees to regularly update their expertise directory of “who knows what” (Wegner, 1995). Given these findings from existing transactive memory research, as well as findings from our study, we argue that field memberships are more valuable than retreats and seed grants for fostering CC collaboration. The reason is that people with multiple field memberships typically have more opportunities to learn about others' expertise when they attend more field faculty meetings (Note: Most fields at Cornell have meetings once or twice per year). Furthermore, people with multiple field memberships also receive more email updates about what is happening in each field in which he or she is involved (Note: All fields at Cornell have mailing lists to distribute information as needed). In summary, our results suggest that whereas periodical institutional interventions in the form of retreats or seed grants are valuable to foster CC collaboration, more permanent institutional arrangements (e.g., multiple field memberships) are even more desirable because they provide people regular opportunities to stay connected and up-to-date with experts from other disciplines.

Second, citation impact evidence presented here suggests that CC publications tended to be cited at a higher rate than SC publications. Although more research is needed to determine the exact reasons for this and the nature of these citation patterns, it does suggest that there is utility to investing in interdisciplinary research. It also suggests that, despite documented coordination difficulties in some multisite collaborations (e.g., Cummings et al., 2007; Nomura et al., 2008), it is possible in at least some cases to overcome these difficulties and produce high-impact research.

Third, a striking feature of these results is the extent to which a very small number of individuals are responsible for nearly half of the CC publication activity. This suggests the importance of individual interest in collaborations, and our informal discussions lead us to believe these individuals are highly motivated to collaborate. Their motivation stems from, for example, interest in interdisciplinary problems (e.g., Birnholtz et al., 2012; Hagstrom, 1965), past success in collaboration (e.g., Cummings & Kiesler, 2008). This suggests that there may be substantial utility in investing resources, such as seed funding or the organization of focused retreats and meetings, in highly motivated individuals, and encouraging them to broaden the network of CC collaborators.

## Limitations and Future Work

There are many limitations to consider in interpreting these results. First, publication data provide us with a clear indicator of who collaborated with whom over the 10 year period of interest, and allow for systematic measurement of factors such as campus affiliation and citation impact. One drawback of this approach, however, is that this provides no record of collaborations that were attempted but failed, or those that simply have not yet generated publishable results. As a general measure of collaboration activity, however, we believe we are able to present useful analyses.

Second, this is a case study examining CC collaboration at a single institution. Although we believe it is possible to derive lessons from this case, it is not possible to discern what elements of these results are attributable to the particulars of the Cornell environment, as opposed to factors that are likely to be present at any institution. This liability is common to all case studies, however, and is not debilitating.

We have combined this approach with complementary qualitative studies to better understand the nature of researchers' collaboration experience (e.g., Birnholtz et al., 2012; Nomura et al., 2008). In future work, we plan to conduct scientometric analyses of future collaborations, specifically examining how collaboration activity varies in frequency and impact over time, and better understanding the role and impact of institutional efforts to facilitate collaboration, including infrastructural elements such as transportation and communication facilities. We plan to augment these analyses with qualitative case studies of specific collaborations and departments, focusing on both CC and SC collaborators to understand individual researcher motives for becoming involved—or not—in collaborative research activities.

## Acknowledgments

This work was supported in part by the National Science Foundation (IIS #0942659 and IOS #0822784). We are also grateful to Richard Wang and Xuan Zhou for their assistance with this research.

## Footnotes

- 1 Note that  $N$  here is higher than article or author totals because some authors appear on more than one article, and all articles necessarily have more than one author.
- 2 Graduate fields are a unique administrative feature of Cornell intended to encourage faculty interaction and collaboration across departmental lines, so it stands to reason that these would impact collaboration trends. Faculty may supervise graduate students in any field of which they are a member, and may belong to multiple fields. Field memberships are independent of departmental affiliation and awarded based on faculty interest and a vote by members of the field. For more information, see: <http://www.gradschool.cornell.edu/academics/fields-study>



3 As the centrality (dependent) variables are continuous, heavily right skewed and positive, a gamma model was selected, based on Culpepper's (2010) recommendations. The equation for the gamma distribution function is given by where  $x \geq 0$  and  $k, \vartheta \geq 0$ .

## References

- Beaver, D. (2001). Reflections on scientific collaboration (and its study): Past, present, and future. *Scientometrics*, 52(3), 365–377.
- Birnholtz, J.P. (2006). What does it mean to be an author? The intersection of credit, contribution and collaboration in science. *Journal of the American Society for Information Science and Technology*, 57(13), 1758–1770.
- Birnholtz, J.P. (2007). When do researchers collaborate? Toward a model of collaboration propensity. *Journal of the American Society for Information Science and Technology*, 58(14), 2226–2239.
- Birnholtz, J.P., Forlano, L., Yuan, Y.C., Rizzo, J., Liao, K., Gay G., & Heller C. (2012). One university, two campuses: Initiating and sustaining research collaborations between two campuses of a single institution. In Proceedings of the 2012 iConference (pp. 33–40). New York: ACM Press.
- Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92(5), 1170–1182.
- Borgatti, S.P., Everett, M.G., & Freeman, L.C. (2002). UCINET 6 for windows: Software for social network analysis.
- Bozeman, B., & Corley, E. (2004). Scientists' collaboration strategies: Implications for scientific and technical human capital. *Research Policy*, 33, 599–616.
- Burt, R.S. (2001). Structural holes versus network closure as social capital. In N. Lin, K. Cook, & R. Burt (Eds.), *Social capital: Theory and research* (pp. 31–56). New York: Aldine de Gruyter.
- Contractor, N., Zink, D., & Chan, M. (1998). IKNOW: A tool to assist and study the creation, maintenance, and dissolution of knowledge networks. In T. Ishida (Ed.), *Community computing and support systems*, Lecture Notes in Computer Science, 1519/1998, 201–217.
- Crane, D. (1972). *Invisible colleges*. Chicago, IL: University of Chicago Press.
- Culpepper, S.A. (2010). Studying individual differences in predictability with gamma regression and nonlinear multilevel models. *Multivariate Behavioral Research*, 45(1), 153–185.
- Cummings, J., & Kiesler, S. (2008). Who collaborates successfully?: Prior experience reduces collaboration barriers in distributed interdisciplinary research. In Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW) (pp. 437–446). New York: ACM Press.
- Cummings, J., & Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy*, 36(10), 138–152.
- Cummings, J., Finholt, T., Foster, I., Kesselman, C., & Lawrence, K. (2008). *Beyond being there: A blueprint for advancing the design, development and evaluation of virtual organizations*. Arlington, VA: National Science Foundation.
- Disis, M.L., & Slattey, J.T. (2010). The road we must take: Multidisciplinary team science. *Science Translational Medicine*, 2(22), 22cm9.
- Elliott, S. (2010). Survey of author name disambiguation: 2004 to 2010. *Library Philosophy and Practice*, 473(November). Retrieved from <http://digitalcommons.unl.edu/libphilprac/473/>.
- Falk-Krzesinski, H.J., Börner, K., Contractor, N., Fiore, S.M., Hall, K.L., Keyton, J., Spring, B., et al. (2010). Advancing the science of team Science. *Clinical Translational Science*, 3(5), 263–266.
- Freeman, L.C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35–41.
- Galison, P., & Hevly, B. (1992). *Big science: The growth of large-scale research*. Stanford, CA: Stanford University Press.
- Hagstrom, W. (1965). *The scientific community*. New York: Basic Books.

- Hara, N., Solomon, P., Kim, S., & Sonnenwald, D. (2003). An emerging view of scientific collaboration: scientists' perspectives on collaboration and factors that impact collaboration. *Journal of the American Society for Information Science and Technology*, 54(10), 952–965.
- Horn, D., Finholt, T., Birnholtz, J., & Motwani, D. (2004). Six degrees of Jonathan Grudin: A social network analysis of the evolution and impact of CSCW research. In Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW) (pp. 582–591). New York: ACM Press.
- Huerta, M.F., Farber, G.K., Wilder, E.L., Kleinman, D.V., Grady, P.A., Schwartz, D.A., & Tabak, L.A. (2005). NIH roadmap interdisciplinary research initiatives. *PLoS Computational Biology*, 1(6), 1.
- ISI Web of Knowledge. (2011). Retrieved from <http://apps.isiknowledge.com>
- Kennedy, D. (2003). Multiple authors, multiple problems. *Science*, 301, 733.
- Koku, E., & Wellman, B. (2004). Scholarly networks as learning communities: The case of TechNet. In S. Barab, R. Kling, & J.H. Gray (Eds.), *Designing for virtual communities in the service of learning* (pp. 299–337). Cambridge, UK: Cambridge University Press.
- Landry, R., & Amara, N. (1998). The impact of transaction costs on the institutional structuration of collaborative academic research. *Research Policy*, 27(9), 901–913.
- Lee K., Brownstein J.S., Mills R.G., & Kohane I.S. (2010). Does collocation inform the impact of collaboration? *PLoS ONE*, 5(12), e14279.
- Lewis, K., Lange, D., & Gillis, L. (2005). Transactive memory systems, learning, and learning transfer. *Organization Science*, 16(6), 581–600.
- Melin, G. (2000). Pragmatism and self-organization research collaboration on the individual level. *Research Policy*, 29(1), 31–40.
- Nentwich, M., (2003). *Cyberscience: Research in the age of the Internet*. Vienna: Austrian Academy of Sciences.
- Newman, M.E.J. (2001). Who is the best connected scientist? A study of scientific coauthorship networks. *Physics Review E*, 64, 16131.
- Nomura, S., Birnholtz, J., Rieger, O., Gay, G., Leshed, G., & Trumbull, D. (2008). Cutting into collaboration: Understanding coordination in distributed and interdisciplinary medical research. In Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW) (pp. 427–436). New York: ACM Press.
- Oliver A.L. (2004). Biotechnology entrepreneurial scientists and their collaborations. *Research Policy*, 33, 583–597.
- Olson, G.M., Zimmerman, A., & Bos, N. (2008). *Scientific collaboration on the internet*. Cambridge, MA: MIT Press.
- Owen-Smith, J., & Powell W. (2003). The expanding role of university patenting in the life sciences: Assessing the importance of experience and connectivity. *Research Policy*, 32(9), 1695–1711.
- Schwenk, T.L., & Green, L.A. (2006). The Michigan clinical research collaboratory: Following the NIH roadmap to the community. *Annals of Family Medicine*, 4(Suppl 1), S49–S54; discussion S58–S60.
- Shami, N.S., Ehrlich, K., & Millen, D.R. (2008). Pick me! link selection in expertise search results. In Proceedings of the ACM Conference on Human Factors in Computing Systems, CHI (pp. 1089–1092). New York: ACM Press.
- Shrum, W., Genuth, J., & Chompalov, I. (2007). *Structures of scientific collaboration*. Cambridge, MA: MIT Press.
- Velden, T., Haque, A. & Lagoze, C. (2010). A new approach to analyzing patterns of collaboration in co-authorship networks: Mesoscopic analysis and interpretation. *Scientometrics*, 85(1), 219–242.
- VIVO Collaboration. (2010). Retrieved from <http://vivo.cornell.edu>
- Wegner, D.M. (1995). A computer network model of human transactive memory. *Social Cognition*, 13(3), 319–339.
- Whitley, R. (2000). *The intellectual and social organization of the sciences*. Oxford: Oxford University Press.
- Who'd want to work in a team? [Editorial]. (2003, July 3) *Nature*, 424, 1. doi:[10.1038/424001a](https://doi.org/10.1038/424001a)

- Wuchty, S., Jones, B.F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–1039.
- Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology*, 60(10), 2107–2118.
- Yuan, Y.C. (2009). Social capital and the development of group transactive memory systems. Saarbrücken, Germany: VDM Verlag Dr. Müller.
- Yuan, Y.C., Carboni, I., & Ehrlich, K. (2010). The impact of awareness and accessibility on expertise retrieval: A multilevel network perspective. *Journal of the American Society for Information Science and Technology*, 61(4), 700–714.
- Yuan, Y.C., Fulk, J., Monge, P.R., & Contractor, N. (2010). Expertise directory development, shared task-interdependence, and strength of communication network ties as multilevel predictors of expertise exchange in transactive memory work groups. *Communication Research*, 37(1), 20–47.