

# **Understanding and enabling network dynamics in virtual communities**

**Research Proposal Funded by the National Science Foundation**

**Noshir S. Contractor, Northwestern University**

**Luís A. N. Amaral, Northwestern University**

**Brian D. Uzzi, Northwestern University**

**Peter R. Monge, University of Southern California**

## Section C – Project Description

### Understanding and enabling network dynamics in virtual communities

#### A. OVERVIEW

To solve the most critical intellectual and social problems that confront us today, we need teams. A central challenge, spurred by new developments in cyberinfrastructure, is that the nature of teams and how they are assembled has changed radically. Today, many teams are ad hoc, agile, distributed, transient entities that emerge from a larger primordial network of relationships within virtual communities. Moreover, teams emerge from multidimensional networks, which include a variety of links that exist not only among individuals, but also with documents, datasets, workflows, analytic tools, and concepts. Hence, all teams are enabled and constrained by the multidimensional networks in which they are embedded. Research conducted by PIs on this project has demonstrated that such team collaborations, a growing trend across all disciplines, yield publications with higher intellectual impact than single researchers. And yet, assembling teams that produce significant knowledge is a daunting task, for intellectual as well as logistic and technical reasons. While there is growing awareness of the socio-economic consequences of team collaborations, our socio-technical understanding of how collaborations emerge and their impact on effectiveness is severely limited.

This project seeks to address this limitation. We propose to develop a theoretical framework, grounded in multidimensional networks, to understand the socio-technical dynamics shaping the assembly of teams in virtual communities. This project addresses two main research questions: First, what are the socio-technical motivations that explain the assembly of teams? Second, to what extent do the assembly mechanisms of teams influence their effectiveness? While there is a vibrant body of related research focused on developing theories and tools to enable and understand how teams collaborate, the current effort is more specifically focused on understanding a precursor to the collaboration itself - the mechanisms by which collaborations are assembled.

Empirically testing such theoretical models pose formidable data collection challenges. However, a unique resource available to the research team is access to six major cyberinfrastructure initiatives serving diverse scientific virtual communities: *nanoHUB*, an NSF funded project serving the nanoscience and technology community; *WATERS Cybercollaboratory*, an NSF funded project serving the hydrologic science and environmental engineering research communities; *NEESit*, an NSF funded project serving the earthquake engineering community; *MyExperiment*, funded by the Virtual Research Environment program in the UK, serving a wide range of basic science communities reuse; *Media Research Hub*, funded by the Ford Foundation and MacArthur Foundations, serving the media research and practitioner communities, and the *Nicotine and Tobacco Research Hub*, funded by the Society for Research on Nicotine and Tobacco, serving the tobacco research community.

Over the past three years, as part of prior and ongoing NSF funded projects, the PIs on this proposal have built partnerships with the leaders of each of these communities. As a result, we have set up mechanisms, with privacy safeguards, to instrument their cyberenvironments, thereby giving us access to all the content as well as behavioral traces (server logs). Members of the team have developed the Cyberinfrastructure for Inquiring Knowledge Networks on the Web (CIKNOW). This program enables us to collate these data with additional relevant data available to the PIs, including

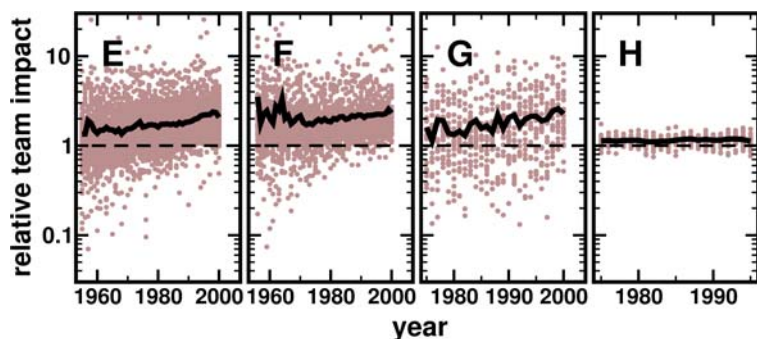
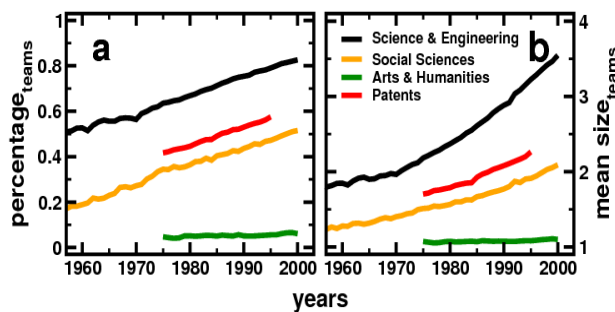
bibliographic data (the entire Web of Science), funding data (from NSF, NIH as well as all publicly funding Foundations in the US), as well as patent data. Taken together, these data offer an unprecedented opportunity to theorize and empirically model the dynamics of the multidimensional networks that shape the assembly of teams in virtual communities. Consequently, this proposal seeks NSF-VOSS funding for a major inter-disciplinary research effort, involving PIs from organizational science, communication, and engineering.

## B. BACKGROUND

### B.1 Emergence of Team Science

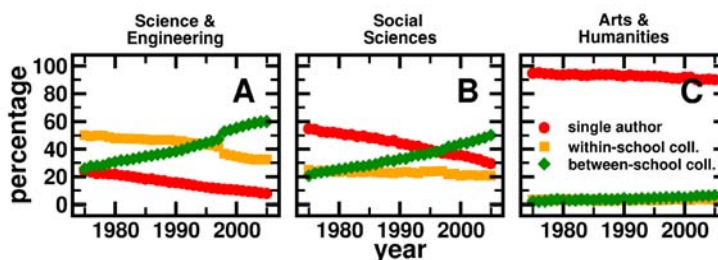
The figure of the lone genius has been the typical model for scientific discovery across the sociology of science and popular culture. (Bowler & Morris, 2005; Merton, 1968). And yet, today we know that this image is more fictional than real. For example, tracing the history of key innovations in art, science, and politics in the ancient Western and Eastern worlds, Collins (1998) showed that only 1<sup>st</sup> century Confucian metaphysicist Wang Ch'ung, 14<sup>th</sup> century Zen spiritualist Bassui Tokusho, and 14<sup>th</sup> century Arabic philosopher Ibn Khaldun fit the pure loner model. Everyone else – including Beethoven, Hutchinson, Hume, Smith, de Medici, and Darwin-- were embedded in self-organizing communities of collaborative relationships among scientists and inventors who shared ideas and acted as both critics and fans for one other (see also Collins, 1998; de Solla Price, 1963).

Turning from case studies to a quantitative analysis, Uzzi, a PI on this proposal, along with colleagues (Wuchty, Jones, and Uzzi, 2007a, 2007b), examined the universality of the shift from the solo to the community model of science, the growth rate of this shift, and its association with high impact research. They studied 19.9 million research articles over 5 decades as recorded in the Web of Science database, and an additional 2.1 million patent records (Hall, Jaffe, & Tratjenberg, 2001) from 1975-2005.



They found three important facts. First, for virtually all fields, research is increasingly done in teams (see Fig above, left). Second, teams typically produce more highly cited research than individuals do (accounting for self-citations), and this team advantage is increasing over time (see Fig above, right). Third, teams now produce the exceptionally high impact research, even where that distinction was once the domain of solo authors (see Fig above).

More recently, they (Jones, Wuchty, Uzzi, 2008) showed that the trend toward virtual communities was not driven by a growth in teamwork by scientists working with other co-located scientists. Rather, the dramatic



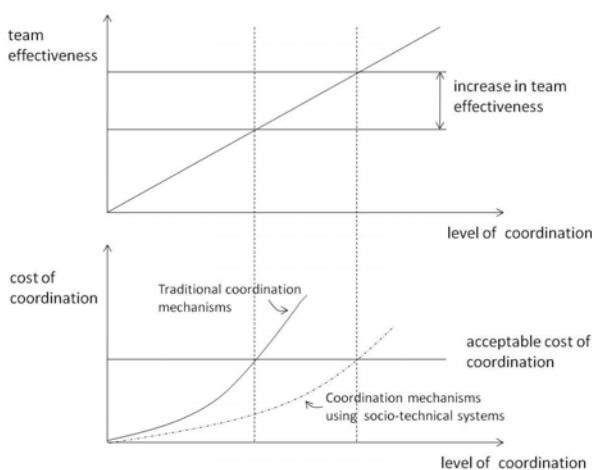
increase in collaboration was due to collaborations that cross university boundaries. Using the Web of Science database to analyze the collaboration arrangements of over 4,000,000 papers over a 30 year period, they found that team science is increasingly composed of co-authors located at different universities. They also found that these “virtual communities of scholars” produce higher impact work than comparable co-located teams or solo scientists. This change is true for all fields and team sizes, as well as for research done at elite universities.

Alongside these trends showcasing the positive impacts of team science, there is research that challenges the superiority of multidisciplinary and multi-university collaborations. Based on a study of two recent NSF initiatives, Knowledge and Distributed Intelligence (KDI) and Information Technology Research (ITR), Cummings and Kiesler (2005, 2007) found that teams involving multiple institutions produced significantly fewer patents, publications and other knowledge outcomes, especially if the team was multidisciplinary. These findings raise new questions of whether teams produce better science and inventions. Teams may bring greater collective knowledge and effort, but they are known to experience social network and coordination losses that make them under-perform individuals even in highly complex tasks. As F. Scott Fitzgerald concisely observed, “no grand idea was ever born in a conference.”

## B.2 Cost of Coordination and Team Assembly

The counterintuitive findings about the effectiveness of team science are ascribed to the tension that exists between the benefits and costs of coordinating (Malone & Crowston, 1994; Van de Ven, Delbecq, & Koenig, 1976) across universities and disciplines. This tension is well captured by Coordination theory (Malone & Crowston, 1994; Malone and Rockart, 1991). As the upper pane in Figure 1 shows, an increase in coordination will lead to greater team effectiveness. However, as the lower pane, indicates, an increase in coordination is accompanied by an exponential increase in the cost of coordination. For instance, it takes alignment of three pairs of calendars to schedule a meeting of three people as opposed to five times as many ( $15 = 6 \cdot 5/2$ ) calendars for a meeting involving twice as many (six) people. Most teams implicitly (or explicitly) agree on an acceptable cost of coordination. This agreement dictates their level of coordination, which in turn influences their team effectiveness.

One promise of new technology is to reduce the cost associated with coordination (shown in the lower pane of Figure 1 with a dotted line). Using technology can enable a higher level of coordination for an acceptable cost, thereby resulting in greater team effectiveness. Currently, most of the literature examines coordination efforts among teams *after they have been assembled*. Examples studied include mechanisms to overcome distance (Hagstron, 1964; Hobday, 2000; Kiesler & Cummings, 2002), to negotiate across disciplines (Metzger & Zare, 1999), to maintain mutual awareness (Weisband, 2002), as well to manage institutional hurdles. Indeed, Cummings and Kiesler (2005, pp. 718-719)



**Figure 1. The role of socio-technical systems in identifying several technological mechanisms that might mitigate the challenges confronting virtual teams: “tools to manage and track the trajectory of**

tasks over time; tools to reduce information overload; tools for ongoing conversation; tools for awareness with reasonable interruption for spontaneous talk; tools to support simultaneous group decision-making; tools to schedule presentations and meetings across distance.”

While studying these technological innovations presents opportunities to reduce the cost already assembled teams incur for coordinating, this line of research does not consider the potential that socio-technical mechanisms can offer to reduce a more primordial coordination cost: the cost of identifying a team’s best members. Research conducted by PIs (Uzzi and Amaral) on our team show that despite the trend toward more frequent and more impactful virtual collaboration, the success of the community may depend on key “assembly” processes. Guimera, Uzzi, Spiro, and Amaral (2005) examined the networks of scientists who collaborate with colleagues at their home university and with colleagues at other universities. Their work showed that there was a direct association between scientists’ connections and the impact of their work. Looking at the co-authorship networks bound within a journal for the top 6-8 journals in four fields (astronomy, ecology, economics, and social psychology), their work showed that the quality of research in these journals varied with the frequency with which the authors in those journals worked repeated with past co-authors or with social distant (in terms of degrees of network separation) co-authors. They found that the best journals had relatively low rates of repeated relationships and high rates of social distance. The cost of coordination incurred in identifying potential co-authors that might be a few degrees of separation apart explains, at last in part, the lack of uniformly impactful collaborations.

As discussed above, and illustrated in Figure 1, this project seeks to identify the socio-technical mechanisms that reduce the cost of coordination associated with the assembly of teams. The following section overviews the potential of Virtual Organization to provide these socio-technical mechanisms.

### **B.3 Definition of VO and its operationalization as a Multidimensional network**

In a special joint issue of *Organization Science* and the *Journal of Computer Mediated Communication*, co-edited by Monge, one of the PIs, and titled “Communication Processes for Virtual Organizations,” DeSanctis and Monge (1999) define virtual organizations (VOs) as “a collection of geographically distributed, functionally and/or culturally diverse entities that are linked by electronic forms of communication and rely on lateral, dynamic relationships for coordination.” (p. 693). Historically, physical location was a major determinant of what organizations could accomplish (McNeill & McNeill, 2003). The Roman Empire, Catholic Church and the Hudson Bay Company were all effective distributed organizations, though they operated at a much slower pace than current organizations (O’Leary, Orlikowski, & Yates 2002).

The term VO was first appropriated by the Grid computing community to describe “flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources — what is referred to as virtual organizations.” (Foster, Kesselman & Tuecke, 1999, p. 200). VO was defined primarily by the use of technologies to do a different and more effective form of science. As the term has evolved over the past decade, the resources being shared include computing cycles, data storage, applications, instruments, data, analytic tools, documents, and workflows (Buetow, 2005; Foster, 2005 ; Hey & Trefethen, 2005).

While this vision preserves the spirit of the original definition of VO offered by DeSanctis and Monge (1999), the advent of VO holds the technological promise of enabling individuals to

seamlessly (i) collaborate with other individuals located anywhere around the globe, (ii) create, share, retrieve or review data, documents, analytic tools, instruments or workflows, located anywhere, and (iii) execute programs using computing cycles residing on computers anywhere. However, the same suite of technologies that enable these tremendous capabilities can also simultaneously impede progress in several ways. They can, for example, hinder researchers' ability to locate other researchers with similar (or complementary) expertise; they can obscure data sets most relevant to a research project; they can confound the tools that would be most appropriate for analyzing data, and they can cloud related concepts that would prove fruitful for gaining new insight into a problem. These challenges increase the cost of coordination necessary for researchers to assemble the most effective teams.

As more resources—across disciplines and distances-- become available to those working within VOs, a major challenge is to help individuals “discover” the most appropriate human and digital resources for the task at hand. Indeed, the cyberinfrastructure community has been making great strides in developing technological mechanisms to facilitate coordination of discovery across this diverse set of resources. These include ontologies, web services, service oriented architecture, Web 2.0 technologies, and semantic grid technologies (Liu, Myers, Minsker, & Futrelle, 2007; De Roure & Goble, 2007). However, by comparison, little is known about individual's social motivations for availing the coordination mechanisms offered by these technological affordances. That is, given a set of individuals embedded in a network of digital resources, what are the social incentives and hurdles that enable them to assemble the most potentially effective team?

How we frame this question emphasizes the need to conceptualize scientists' social networks in a new way: as multidimensional networks where nodes are “resources” --including people, documents, datasets, analytic tools, or instruments. Traditionally, social network analysis has investigated “single mode” networks – such as people-to-people, or bimodal affiliation networks-- such as connections between people and the events they attend (Wasserman & Faust, 1994). As part of our prior and ongoing NSF-funded research, we extended traditional approaches to social network analysis and helped develop the theoretical and methodological apparatus to conceptualize and analyze communities as multidimensional networks in terms of multiple types of nodes and relations (Contractor, 2002, 2005; Contractor, Wasserman, and Faust, 2006; Monge & Contractor, 2003).

A very large number of relations can exist within multidimensional networks. Consider first the relations that exist between *people* and the some of the nodes within these multidimensional networks:

- (i) *people to people* (such as collaboration, co authorship, citation);
- (ii) *people to document* (people authoring, publishing, citing, retrieving, credentialing or rating a document );
- (iii) *people to data sets* (people publishing, retrieving, credentialing or rating data sets);
- (iv) *people to tools* (people developing, utilizing, or credentialing/rating these tools);
- (v) *people to keywords/concepts* (people who have published widely on certain keywords/concepts).

In addition to relations between people and other nodes, we can also consider relations among and between the remaining types of nodes. For instance:

- (i) *document to data sets* (documents that report on specific data sets);
- (ii) *data sets to visualization - analytic tools* (data sets that are visualized-analyzed using specific tools);
- (iii) *data sets to keywords* (data sets that contained data about specific keywords or concepts);

- (iv) *documents to documents* (those who created/cited/retrieved/rated one document created/cited/retrieved/rated the other document);

## C. RESEARCH OBJECTIVES

Given the coupling of the trend toward team science with the unprecedented rise in the importance of virtual organizations, we propose to investigate several key questions about their functioning, with an eye toward the factors that may promote or stifle the community's success in producing valuable knowledge. Broadly, the proposed study addresses two research questions:

*Research Question 1.* What are the socio-technical assembly mechanisms for collaborations within virtual communities? How are individuals' motivations for assembly enabled and constrained by the multidimensional networks in which they are embedded?

*Research Question 2.* How do the assembly mechanisms of teams influence their ability to be more effective in producing new knowledge (innovation in disciplinary, multi-/inter-/trans-disciplinary research; patents, tools, workflows, instruments, data; graduate and undergraduate training, outreach with industry and community)?

### C.1 Socio-technical assembly mechanisms for collaborations

In prior and ongoing NSF-funded research, PIs on our team have developed a multi-theoretical multilevel (MTML) model to explain individuals' motivations to create, maintain, dissolve, and reconstitute links with others in a network (Contractor & Monge, 2002; Contractor & Monge, 2003; Contractor, Wasserman & Faust, 2006; Monge & Contractor, 2003). The model explains these motivations on the basis of attributes of the individuals (such as role, expertise) as well as the links (such as collaboration, citation) among individuals within the network.

Since our research question here is to understand the mechanism of team assembly, the proposed effort extends our prior research by specifically addressing individuals' motivations to create, maintain, dissolve, or reconstitute a *team* linkage with another individual (Katz, Lazer, Arrow, & Contractor, 2004). In addition to examining linkages among individuals, we also take into account the multidimensional networks in which these individuals are embedded. For instance, we consider to what extent a collaboration link between two individuals is motivated by, say, one creating (and thus having a link to) a dataset, which can be analyzed using a tool created by another individual? Next, we discuss the various MTML mechanisms illustrating their explanation for creating team assembly linkages.

The MTML model posits eight families of team assembly mechanisms:

(1) *Theories of self-interest* focus on how people make choices that favor their personal preferences and desires, creating team ties that enable them to maximize the goals they wish to achieve. For instance, individuals in virtual communities will join a team that has members who have access to datasets or analytics that they do not possess but from which they seek to benefit. Two primary theories in this area are the theory of social capital (Burt, 1992) and transaction cost economics (Williamson, 1991).

(2) *Theories of mutual interest and collective action* examine how forging links produces collective outcomes unattainable by individual action. Individuals create collaboration ties because they

believe they serve their mutual interests in accomplishing common or complementary goals (Fulk et al., 2004). For instance, scientists might contribute their individual data to community data systems that can then be analyzed by the community.

(3) *Contagion theories* address questions pertaining to the spread of ideas, messages, attitudes, and beliefs through direct or indirect groups links (Burt, 1987). Scientists may seek to join a team simply because colleagues who use the same instrument as they do are joining the team and have infected them with the idea. On the other hand, collaboration can be blocked by isolating parts of the network or by inoculating against infection.

(4) *Cognitive theories* explore the role that meaning, knowledge, and perceptions play in development of teams. Grounded in Transactive memory, decisions to team with others are influenced by what people know (Hollingshead, Fulk, & Monge, 2002). Research has argued that increasing specialization drives collaboration, which, coupled with the limited capacity of organizations to encapsulate more than a fraction of a field's specializations, suggests that specialization in a field may promote collaborations as collaborators search ideas that can be combined to create high impact science and invention.

(5) *Exchange and dependency theories* explain the emergence of teams on the basis of the distribution of information and material resources among network members (Cook, 1982). People seek group ties with those whose resources they need and who in turn seek resources they possess. Individuals who are data analysts might seek collaborations with data providers so that collectively they can exchange resources they need with resources they can offer.

(6) *Homophily and proximity theories* account for emergence of group links on the basis of trait similarity and similarity of place (McPherson & Smith-Lovin, 1987; McPherson, Smith-Lovin, & Cook, 2001). Jones, Wuchty, and Uzzi (2008) showed that collaborators who were part of multi-university teams had a strong preference for choosing co-authors at universities with a similar rank. Scientists at top tier universities became increasingly likely to work with top tier and increasingly less likely to work with co-authors from lower tier universities. This suggests that matches among collaborators in virtual communities may be based less on geographic proximity and more on social, in-group matching based on status homophily.

(7) *Balance theories* (Heider, 1958; Holland & Leinhardt, 1975) posit a consistency towards relations. That is, individuals are more likely to create transitive ties. For instance, scientists are more likely to join teams with those who cite the same documents they do.

(8) Finally, *coevolutionary theory* posits that group linkages are typically created in the belief that they will increase group fitness, measured as performance, survivability, adaptability, and robustness (Campbell, 1986). Coevolutionary theory articulates how communities linked by intra- and inter-group networks compete and cooperate with each other for scarce resources (Baum, 1999; Bryant & Monge, 2008; Monge, Heiss & Margolin, in press). For instance, members of one team assembled on the basis of homophily might observe that they are less effective than a competing team which used, say, an assembly mechanism based on exchange.

Clearly, not all of these mechanisms will operate in all contexts and at all times. A major focus of this project, then, is to develop a contingency framework identifying the conditions under which one or some limited ensemble of assembly mechanisms are more influential than others. For instance, based on a review of 212 laboratories, Bos et al (2007, p. 17) distinguished between six different types of laboratories that they arrayed along two dimensions: resource and activity. The resource dimension was whether the focus of the laboratory was on tools (instruments), information (data) or knowledge (new findings). The activity dimension was whether the focus of the

collaboratory was on aggregating across distance (loose coupling, often asynchronously) or co-creating across distance (requiring tighter coupling, often synchronously). We will explore the extent to which this framework might serve to distinguish distinct ensembles of team assembly mechanisms.

In summary, this section has proposed a series of theoretical mechanisms that influence team assembly. The mechanisms take into account not only the ties among the individuals but also the ties within the multidimensional networks in which these individuals are embedded. The contingency framework proposes that the likelihood of an ensemble of theoretical mechanisms explaining team assembly will depend on the resources and activity of the specific virtual community.

## **C.2 The impact of assembly mechanisms on team effectiveness**

The previous section discussed the socio-technical mechanisms that influence the assembly of teams. However, as discussed previously, there is considerable variance in the effectiveness of teams. We define team effectiveness in terms of generating the following: (i) new knowledge and inventions, including publications and patents; (ii) visual analytics tools, simulation tools, remote instruments, and workflows; (iii) digital repositories for structured or unstructured data and knowledge, (iv) formal and informal training of undergraduate and graduate students via learning modules, curriculum development, and (v) outreach partnerships with industry as well as community organizations such as schools and museums.

This definitional framework needs to be implemented to be useful. To develop indicators for the quality of published research, citation analysis has been both equally criticized (MacRoberts & MacRoberts, 1989) and championed (Baldi, 1998; Leydesdorff & Wouters, 1999; Nicolaisen, 2007). One of the PIs, Amaral, recently investigated two fundamental aspects concerning the prediction of the ultimate impact of a published research paper (Stringer, Sales-Pardo, & Amaral, 2008): (i) the time scale  $\tau$  for the full impact of papers published in a given journal to become apparent, and (ii) the typical impact of papers published in a given journal. This research found that  $\tau$  varies from less than 1 year to nearly 26 years, depending on the journal. They additionally found that there is a typical value and a well-defined range for the eventual impact of papers published in a given journal. This finding enabled them to develop a model for the distribution of article impact that is in good agreement with the data (Stringer, Sales-Pardo, & Amaral, 2008). We will deploy these models to assess, more accurately than was heretofore possible, the long term impact of the effectiveness of collaborations.

As indicated earlier, the preponderance of research on collaboration so far has sought to explain variance in outcomes in two main ways: by focusing on a team's composition and demographics (distance and discipline) (Cummings & Kiesler, 2005; 2007), and by examining the socio-technical mechanisms enabling collaboration processes (e.g., Finholt, 2002; Finholt & Olson, 1997; Kraut, Egido & Galegher, 1990). Until recently, scant research has examined how team assembly mechanisms and underlying network dynamics impact collaboration outcomes. A recent study by Uzzi, a PI on this project (Jones, Wuchty, & Uzzi, 2008) found that scientists who partner with scientists at other top schools garner notable gains in frequency of citations to their work. However, more surprising was the finding that there is relatively little or no loss for partnering with lower-tier schools. This finding suggests that effective collaborations may follow a "strongest-link"

rather than a “weakest-link” model of teamwork, with the strongest (high tier school) link trumping the weakest (low tier school) link. At the institutional level, these findings suggest that whereas the greater geographic interconnectedness of universities through virtual collaborations of their faculty would appear to make geography less important, the corresponding intensification of social stratification in multi-university collaboration will tend to embed the production of outstanding scientific knowledge in fewer rather than more centers of high impact science.

In addition to understanding the overall production of knowledge, there is growing policy interest to assess the extent to which the assembly mechanisms of teams are advancing the production of disciplinary, multi-disciplinary, inter-disciplinary or trans-disciplinary knowledge (Stokols, et al. 2003, 2005). Multi-disciplinary research is defined as collaborations across disciplines that advance each of the disciplines independently. Inter-disciplinary research is defined as collaborations across disciplines where concepts from one discipline are incorporated into another discipline. Trans-disciplinary research is defined as collaborations across disciplines where new concepts emerge that transcend the collaborating disciplines. Recent advances in text mining make it possible to measure the extent of M/I/T-disciplinarity in publications emerging from research collaboration.

Based on the assumption that the resources for understanding and resolving many of the grand societal challenges are not confined to one discipline, several government funding agencies and private foundations are making substantial investments in these M/I/T-disciplinary research initiatives. The hypotheses that these investments are yielding the desired M/I/T-disciplinary objectives remain largely untested. From a theoretical standpoint, we will explore how the team assembly mechanisms influence the likelihood of effectiveness M/I/T-disciplinary research.

## D. METHODS

### D1. Test Beds

A unique resource available to the research team is access to six major cyberinfrastructure initiatives serving diverse scientific virtual communities:

1. *nanoHUB*, (<http://www.nanohub.org/>), an NSF funded project serving the nanoscience and technology community;
2. *WATERS*, (<http://www.watersnet.org/>), an NSF funded project serving the hydrologic science and environmental engineering research communities;
3. *NEESit*, (<http://it.nees.org/index.php>), an NSF funded project serving the earthquake engineering community;
4. *MyExperiment*, (<http://www.myexperiment.org/>), funded by the Virtual Research Environment program in the UK, helping a wide range of basic science communities reuse and share workflows;
5. *Media Research Hub*, (<http://mediaresearchhub.ssrc.org/>), funded by the Ford Foundation and MacArthur Foundations, serving the media research and practitioner communities; and the
6. *Nicotine and Tobacco Research Hub*, (<http://www.srnt.org/>), funded by the Society for Research on Nicotine and Tobacco, serving the tobacco research community.

Members of these initiatives range in disciplines from engineering (*NEES*, *WATERS*, *nanoHUB*), basic sciences (*myExperiment*), life sciences (*Nicotine and Tobacco Research Hub*) to social sciences

(*Media Research Hub*). They vary in longevity from *NEES* (which is close to a decade old) to *Nicotine and Tobacco Research Hub* (that is just underway). Further, they include all the six functionalities of VOs identified by Bos, et. al (2007): shared instruments, community data systems, virtual communities of practice and learning, community infrastructure, open community contribution systems, and distributed research centers.

The letters of commitment included with this proposal provide details about the history, size, activities and goals of each virtual community. The letters also describe their ongoing and future commitments for making content and data available to this project. These letters of cooperation attest to long-standing relationships we have developed with each of these partners.

## **D2. Description of CIKNOW Technical architecture**

The study will leverage the NSF-funded CIKNOW (Cyberinfrastructure for Inquiring Knowledge Networks on the Web) infrastructure to capture, collate, analyze, and visualize the multidimensional network from each of the six virtual communities (Green, Contractor, & Yao, 2006; Huang, Contractor, & Yao, 2008). The CIKNOW infrastructure adopts open-source technologies and software components including CentOS operating system, MySQL databases, and Apache/Tomcat web servers with JK Connectors. The full time research programmer and research scientist who manage and extend the CIKNOW architecture are supported with ongoing funds and are not requested as part of this proposal (see attached letter from Northwestern University).

CIKNOW receives data from two primary sources. First, the six virtual communities each provide activity logs in the form of multidimensional networks, with nodes being digital artifacts available in the virtual environment and links being the activity transactions among these artifacts. Table 1 below provides details of the nodes and the edges for which data are captured from each of the six communities.

Second, based on the individuals identified, CIKNOW uses third party data sources to build publication, funding, and patent databases, thereby providing additional information and relations for the individuals and other digital artifacts in each virtual community. The publication database developed by one of the PIs, Amaral, includes 7.6 million authors and 36 million publications and updates the content through Web of Science academic database (WoS). Using Entrez Programming Utilities (eUtils), CIKNOW also queries PubMed as well as other NCBI databases for biomedical data. The funding databases within CIKNOW are updated periodically from NIH/CRISP, NSF's funding database, and other foundations using data provided by Innolyst's ResearchCrossRoads.com web services. The attached letter of cooperation indicates their commitment to continue providing us with these data at no charge. Patent information is collected from PatFT and AppFT.

The activity logs from the six virtual communities and the third party data sources for publication, funding, and patent databases are collated using the CIKNOW architecture to generate the multidimensional network for each community. The analytic components within CIKNOW also compute the collaboration outcome variables identified.

**Table 1. Details of multidimensional networks sourced from six virtual communities**

Node types	nanoHub	MediaHub	myExperiment	NEESit	WATERS	RNT-Hub
Users	X	X	X	X	X	X
Keywords	X		X	X	X	X
Projects	X			X		
Groups	X		X	X	X	
Publications	X	X	X		X	X
Learning Modules	X					
Analytic Tools	X			X		
Meetings	X					
Workflows			X		X	
Reviews	X		X			
Blogs	X	X	X		X	X
Forum posts,	X		X		X	
Instruments				X		
Organizations				X		
Grants		X		X		
Data files		X		X	X	
News	X	X		X	X	X
Events	X	X		X	X	
Email Lists		X		X	X	

Edge/relation types	nanoHub	MediaHub	myExperiment	NEESit	WATERS	RNT-Hub
Create, upload	X	X	X		X	X
Access, download	X	X		X	X	X
Review, comment	X		X		X	
createTag	X		X			X
hasTag	X		X			X
hasFriend			X		X	
Group	X		X	X	X	
attendMeeting	X					
HasGrant		X				
consistOf	X			X		
Citation	X		X			X

### D3. Analyses

The team assembly mechanisms and their impacts on outcomes will be tested using exponential random graph models, ERGMs (also known as p\*). The multidimensional network data harvested above is represented as a large multigraph with many types of actors and many relations. In addition, we have attribute information on the actors. We assume that these quantities are random variables “tied together” by the theoretical concerns under study. Each of the eight hypothesized team assembly mechanisms has its unique structural signatures (see Contractor, Wasserman, & Faust, 2006 for a review of all structural signatures). For instance, if the assembly was motivated by balance, one would expect to observe more transitive triads than one would expect by chance. The general form of the class of (homogeneous) exponential random graph models is as follows:

$$P(\mathbf{X} = \mathbf{x}) = \kappa^{-1} \exp \left( \sum_{A \subseteq N_D} \lambda_A z_A(\mathbf{x}) \right), \quad (1)$$

where:

- (i) the summation is over structural signatures of types  $A$ ;
- (ii)  $\lambda_A$  is the parameter corresponding to structural signatures of type  $A$ ;
- (iii)  $z_A(\mathbf{x})$  is the *network statistic* corresponding to structural signature  $A$
- (iv)  $\kappa$  is a normalizing quantity to ensure that (1) is a proper probability distribution.

The model represents a probability distribution of graphs on a fixed node set, where the probability of observing a graph depends on the presence of the various structural signatures hypothesized in the model. One can interpret the structure of a typical graph in this distribution as the result of a cumulation of these particular structural signatures. The estimated parameters provide information about the presence of those structural signatures in the observed data (Robins, Snijders, Wang, Handcock, & Pattison, 2007). Extensions to this basic exponential random graph model (ERGM) will be used to address Research Question 1 (hypotheses about links being created or “social selection”) and Research Question 2 (hypotheses about outcomes resulting from links being created or “social influence”).

## E. EDUCATIONAL & DIVERSITY PLAN

A natural educational outcome of a project such as this is a cross-disciplinary graduate level course on Team Science. This course would analyze the antecedents and outcomes of team assembly across communication, organizational, and computational perspectives. The PIs will develop and offer such a course between Northwestern and USC. The pedagogic model will build on the experiences of 7 distributed graduate courses taught by Contractor and colleagues under the auspices of the Committee for Institutional Cooperation (Academic Big Ten). *Second*, we will develop instructional/tutorial modules based on this project for the NetSci International Network Science Conference, at which Contractor has offered tutorials in 2006, 2007, and 2008 and Uzzi is delivering the keynote in 2008. *Third*, the PIs have a track record of working with minority students through the Summer Research Opportunity Program (SROP) and the McNair undergrad summer training programs. In addition, the Science of Networks in Communities (SONIC) lab is hosting a graduate student who has been using the CIKNOW infrastructure to develop virtual communities of mentoring for minority students in Computer Science as part of the NSF funded Empowering Leadership Alliance Mentoring Program. *Fourth*, the creation of such a massive data set will be an immense opportunity for graduate students to pursue their own dissertation and thesis projects. No previous effort will have combined the size and scope of the dataset to be gathered here. *Finally*, all PIs have established a track record for cultivating on-going collaborations among our graduate students across multiple disciplines, in joint research as well as instructional activities. As such, the proposed project will continue to nurture a new generation of graduate scholarship where a premium is placed on collaborative fluency across traditional disciplinary boundaries.

## F. PROJECT ORGANIZATION

This project provides a key opportunity for experienced researchers across a wide range of disciplines to work together on a problem that requires their collective expertise. All the PIs have significant experience in cross-disciplinary research, and a record of working with each other. Amaral and Uzzi have co-authored multiple papers including one published in *Science*; they have co-edited a special issue of the journal *Management Science*. Monge and Contractor have coauthored a

book and several articles over the past two decades. Contractor, Amaral and Uzzi are also co-investigators on a project funded internally by Northwestern University on the Science of Team Science (see attached letter from Northwestern University). The four PIs have agreed to an efficient management plan to deliver the proposed research results. Expertise will be drawn from any of the PIs, but specific PIs will be responsible for certain research tasks and deliverables. As PI, Contractor will coordinate all aspects of the project. The *Data Management & Informatics* effort will be led by Amaral, *Visual-analytics and Statistical Modeling* by Contractor, *Theory Development* by Monge, and *Dissemination* by Uzzi. The project management team will hold biweekly meetings conducted using Internet-based conferencing facilities at our respective universities. In addition, the PIs' labs at Northwestern and USC have access to persistent Internet based videoconferencing facilities.

## G. WORK PLAN AND SCHEDULE

We plan to complete this research project in three years. The detailed schedule of the planned tasks in the areas of data, research, education, and dissemination are shown in Table 2 below.

ID	Task Name	PI	2008				2009				2010				2011		
			4	1	2	3	4	1	2	3	4	1	2	3			
<b>1. Data Tasks</b>																	
1.1	Set up system for data transfer from VOs to CIKNOW	C	■	■	■	■											
1.2	Set up system for data transfer from 3rd party dbases to CIKNOW	A	■	■	■	■											
1.3	Transform Raw Data to Generate Multidimensional Networks	C			■	■	■	■									
1.4	Develop Mechanisms for PI access to Data	M, U					■	■	■	■							
1.5	Provide Ongoing Data Access, Computation Support, etc.	N								■	■	■	■	■	■	■	
<b>2. Research/Analysis Tasks</b>																	
2.1	Identify structural signatures for team assembly mechanisms	C, M			■	■	■										
2.2	Computation of network metrics for team effectiveness	A			■	■	■										
2.3	Extend ERGMs methods for large scale networks	A			■	■	■	■									
2.4	Estimate parameters for team assembly	U						■	■	■	■	■					
2.5	Estimate parameters for effectiveness	U								■	■	■	■				
2.6	Comparative analyses across six test beds	M											■	■			
<b>3. Education &amp; Dissemination tasks</b>																	
3.1	Teach course on team science	C, M								■	■	■					
3.2	Share results with virtual community leaders and members	A, C, M, U								■	■	■	■	■	■	■	
3.3	Present results to Network Science, Team Science, audiences	A, C, M, U											■	■	■	■	

**Table 2. Work plan for research, education and dissemination**

## H. DISSEMINATION AND OUTREACH

The PIs have extensive experience and leadership roles in disseminating their findings in prestigious forums in the areas of organizational science, network science, team science and communication. The four PIs will continue to present materials relevant to this proposal at the following academic venues: AAAS, International Communication Association, Academy of Management, INFORMS, Network Science, the Social Network “Sunbelt conference.” The PIs expect to publish results of our

research in journals where we have published previously including *Science*, *Nature*, *PNAS*, *Harvard Business Review*, *Academy of Management Journal*, *Organization Science*, *Management Science*, *Social Networks*, *PLoS*, *American Journal of Sociology*, and *Journal of Communication*. In addition, the PIs will continue to share insights and progress on our efforts at annual workshops and “all-hands” meetings of the six virtual communities that are partnering on this project.

## **I. INTELLECTUAL MERITS & FIT TO VOSS**

*First*, the proposed research offers the promise to usher in a new generation of theorizing and research on the assembly mechanisms of teams embedded in the multidimensional networks offered by virtual communities. The empirical data that will be used to develop and test these theories will be unprecedented in scale, size, and completeness. *Second*, the project will offer groundbreaking insights into how the network configurations shaping the assembly of teams created in virtual organizations impact their subsequent effectiveness in disciplinary as well as Multi-/Inter-/Trans-disciplinary knowledge production, education, and outreach. *Third*, it will contribute to advances in techniques for modeling the effectiveness of teams based on digital traces from scientometric, bibliometric and usage logs from a diverse set of virtual communities. *Fourth*, the findings across the six test beds will provide comparative insights on the assembly mechanisms of teams as well as their influence on effectiveness, level of maturity, and size of virtual communities across disciplines. *Finally*, our research will significantly extend exponential random graph modeling techniques. Previous techniques have been used to estimate structural dependencies in 10,000-node networks. Our technique will analyze networks of over a million nodes.

## **J. BROADER IMPACTS & FIT TO VOSS**

The proposed research will have broad impacts on at least three stakeholder communities. *First*, the increasing numbers of researchers who are joining virtual communities need to navigate the opportunities and challenges that these virtual environments present. The results of our study have a direct import on their efforts to understand how to assemble effective teams. Our results will be particularly relevant to facilitate inclusiveness for researchers from non-elite institutions, given our preliminary findings that suggest a tendency for researchers at elite institutions to assemble in teams with others at elite institutions, even though the outcomes are no more effective than if they partnered with a researcher at a non-elite institution. *Second*, the findings of our study are also of interest to leaders of large virtual organizations who view the visual-analytics derived from our multidimensional network as a cyber-dashboard to assess the health of their virtual community at the individual, team, and community levels; and to steer strategies in response. The strength of our commitment letters from the leaders of these communities testifies to the benefits they derive from this partnership. *Third*, funding agencies and science policy makers have a special interest in viewing the ecology of virtual communities as part of their portfolio assessment. The proposed research provides theoretical and methodological insights on how the socio-technical mechanisms for assembling teams influence the intellectual, human resource, and monetary returns on their investments in large scale funding initiatives. Even prior to making an investment, the theories and methodologies developed in this research will provide them with a set of diagnostics to assess the Cyberinfrastructure readiness (CI-readiness) of a community.

## K. SUMMARY OF PRIOR NSF SUPPORT

**Contractor** has received continuous NSF funding for the past 14 years. In the past five years, he has been PI on **OCI-0753047** (\$199,619; 2/15/2008-1/31/2009): *CI-KNOW Cyberinfrastructure Tools to Enable Knowledge Network Discovery, Diagnosis and Design*; PI on **IIS-0729505** (\$604,755; 9/1/2007-8/31-2010): *Virtual Worlds: An Exploratorium for Theorizing and Modeling the Dynamics of Group Behavior*; co-PI on **IIS-0628036** (\$155,851, 4/06-4/07): *Instrumenting behaviors and attitudes in virtual worlds*; PI on **IIS-0535214** (\$808,089; 9/05-9/08): *Social Networking Tools to Enable Collaboration in the Tobacco Surveillance, Epidemiology, and Evaluation Network (TSEEN)*; PI on **SBE-0555115** (\$140,000, 2/06-2/07): *Collaborative Research: Mapping and Analyzing Emergent Multiorganizational networks in the Hurricane Katrina Response*; co-PI on **ITR CMS-0427089** (\$2,370,000.; 9/04-9/09): *IT-Based Collaboration Framework for Preparing Against, Responding to, and Recovering from Disasters Involving Critical Physical Infrastructures*. The research in these projects has resulted in (i) a novel context-driven Multi-Theoretical MultiLevel (MTML) model for the emergence of networks, (ii) extending Exponential Random Graph modeling techniques to test MTML models; (iii) the development of a suite of visual-analytic and navigation tools for social networks: *CI-KNOW* (Cyberinfrastructure for Knowledge Networks On the Web), (iv) the development of *Blanche*, an agent based computational network modeling environment. The findings from these projects have been disseminated in an award winning book titled *Theories Of Communication Networks* (Oxford, 2003 and coauthored with Peter Monge), two dozen articles and chapters, a dozen keynote lectures, and over 60 presentations at conference/workshops in 12 countries. The projects contributed to the training of 5 Post docs, 6 international visiting scholars, 18 graduate students (8 females) and 9 PhD. dissertations (5 female).

**Amaral** is a Co-Principal Investigator on SBE 0624318, Exploring Educational Policy and Change from a Complex Systems Perspective. The research in this project has resulted in the creation of new tools based on agent-based modeling and social network analysis that will improve current understanding of school reform initiatives. This research has also resulted in three papers authored by Amaral which were published in *The Proceedings of the National Academy of Sciences*, *Nature Physics*, and *Physical Review E*.

**Monge** has received four NSF grants over the past two decades. Most recently, he was CO-PI of **IIS-9980109** (\$1,500,000; 9/99-8/04): *Co-evolution of Knowledge Networks and 21st Century Organizational Forms: Computational Modeling & Empirical Testing* (UIUC, CMU, Stanford, USC). The research in this project has resulted in 14 published articles in leading organizational and communication journals, another eight book chapters, and more than 20 conference papers. It also supported the writing of an award winning book, *Theories of Communication Networks* (Oxford University Press), coauthored with Noshir Contractor, which developed a Multi-theoretical Multilevel model of network analysis to explain the emergence, transformation, and eventual demise of organizational networks. The project supported dissertations for six doctoral students, and more than a dozen presentations at leading universities (Stanford, MIT, Harvard, Cornell, etc.).

## References

- Aldrich, H.E., and Ruef, M. (2006). *Organizations evolving* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage.
- Baum, J.A.C. (1999). Whole-part coevolutionary competition in organizations. In J.A.C. Baum & B. McKelvey (Eds.), *Variations in organization science: In honor of Donald T. Campbell* (pp. 113-135). Thousand Oaks, CA: Sage.
- Baldi, S. (1998). Normative versus social constructivist processes in the allocation of citations: A network analytic model. *American Sociological Review* 63, 829–846.
- Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., Dahl, E., et. al. (2007). From shared databases to communities of practice: A taxonomy of collaboratories. *Journal of Computer-Mediated Communication*, 12, 2, article 16. <http://jcmc.indiana.edu/vol12/issue2/bos.html>
- Bowler, P. J., & Morus, I. R. (2005). *Making Modern Science: A Historical Survey* Chicago: University of Chicago.
- Bryant, J. A., & Monge, P. (2008). The evolution of the children's television community, 1953-2003. *International Journal of Communication*.
- Buetow, K. H. (2005). Cyberinfrastructure: Empowering a "Third Way" in Biomedical Research Science, 308, 5723, 821 - 824
- Burt, R. S. (1987). Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology*, 92, 1287-1335.
- Burt, R. S. (1992). *Structural holes: The social structure of competition*. Boston: Harvard University Press.
- Campbell, D. T. (1986). An organizational interpretation of evolution. In D. D. Depew & B. H. Weber (Eds.) *Evolution at a crossroads: The new biology and the new philosophy of science* (pp. 33-167). Cambridge, MA MIT Press
- Collins, R. (1998). *The Sociology of Philosophies: A Global Theory of Intellectual Change* Cambridge, MA: Harvard University Press.
- Contractor, N.S. (2002). Introduction: New media and organizing. In L. Lievrouw & S. Livingstone (Eds.), *Handbook of new media* (pp. 201-205). London: Sage.
- Contractor, N (March, 2005). *The role of social network analysis in enabling cyberinfrastructure and the role of cyberinfrastructure in enabling social network analysis*. White paper prepared for the National Science Foundation workshop on Cyberinfrastructure for the Social Sciences, Airlie House, Warrington, Virginia, USA.
- Contractor, N. S., & Monge, P. R. (2002). Managing knowledge networks. *Management Communication Quarterly*, 16, 249-258.

- Contractor, N. S., & Monge, P. R. (2003). Using multi-theoretical multilevel models to study adversarial networks. In R. Breiger, K. Carley, & P. Pattison (Eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers* (pp. 324-344). Washington D.C.: National Research Council.
- Contractor, N., Wasserman, S., & Faust, K. (2006). Testing multi-theoretical multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review*, 31, 681-703.
- Cook, K. S. (1982). Network structures from an exchange perspective. In P. V. Marsden & N. Lin (Eds.), *Social structure and network analysis* (pp. 177-218). Beverly Hills, CA: Sage.
- Cronin, B., Shaw, D., & La Barre, K. (2003). A cast of thousands: co-authorship and sub-authorship collaboration in the twentieth century as manifested in the scholarly literature of psychology and philosophy. *Journal of the American Society for Information Science and Technology*, 54 (9), 855-871.
- Cummings, J., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science*, 35(5), 703-722.
- Cummings, J. N., and Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy*, 36, 1620-1634.
- De Roure, D. & Goble, C. (2007) myExperiment - A Web 2.0 Virtual Research Environment. In: *International Workshop on Virtual Research Environments and Collaborative Work Environments*, Edinburgh, UK. <http://eprints.ecs.soton.ac.uk/13961/>
- de Solla Price, D. J. (1963). *Little Science, Big Science*. New York: Columbia University Press.
- DeSanctis, G., & Monge, P. (1999). Communication processes for virtual organizations. *Organization Science*, 10(6), 693-703.
- Finholt, T. A. (2002). Collaboratories, *Annual Review of Information Science and Technology*, 36, 73-107
- Finholt, T. A., & Olson, G.M. (1997). From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science* 8, 1, 28-36.
- Foster, I (2005). Service-oriented architecture, *Science*, 308, 5723, 814 - 817
- Foster, I., Kesselman, C., & Tuecke, S. (1999). The Anatomy of the Grid: Enabling Scalable Virtual Organizations, *International Journal of High Performance Computing Applications*, 15, 3, 200-222.
- Fulk, J., Heino, R., Flanagan, A., Monge, P., & Bar, F. (2004). A test of the individual action model of organizational information commons. *Organization Science*, 15, 569-585.
- Green, H. D., Contractor, N. S., & Yao, Y. (2006). CI-KNOW: Cyberinfrastructure Knowledge Networks on the Web. A Social Network Enabled Recommender System for Locating Resources in Cyberinfrastructures, *Eos Trans. AGU* 87(52), Fall Meet. Suppl. 2006.

- Guimera, R., Uzzi, B., Spiro, J. & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308, 697–702.
- Hagstrom, W. (1964). Traditional and modern forms of scientific teamwork. *Administrative Science Quarterly*, 9(3), 241-264.
- Hall, B. H., Jaffe, A. B., Tratjenberg, M. (2001). *National Bureau of Economic Research Working Paper #8498*.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: John Wiley.
- Hey, T., & Trefethen, A. E. (2005). Cyberinfrastructure for e-Science, *Science*, 308, 5723, 817 – 821.
- Hobday, M. (2000). The project-based organisation: An ideal form for managing complex products and systems? *Research Policy*, 29, 871-893
- Holland, P.W., & Leinhardt, S. (1975). The statistical analysis of local structure in social networks. In Heise, D. R., (Ed.), *Sociological Methodology, 1976*, 1-45. San Francisco: Jossey Bass.
- Hollingshead, A.B., & Contractor, N. (2002). New media and organizing at the group level. In L. Lievrouw & S. Livingstone (eds.), *The handbook of new media* (pp. 221-235). Thousand Oaks, CA: Sage.
- Hollingshead, A. B., Fulk, J., and Monge, P. (2002). Fostering intranet knowledge-sharing: An integration of transactive memory and public goods approaches. In S. Kiesler & P. Hinds (Eds.), *Distributed Work: New Research on Working across Distance using Technology* (Pp. 235-255). Cambridge, MA: MIT Press.
- Huang, Y., Contractor, N., Yao, Y. (2008). *CI-KNOW: Recommendation based on Social Networks*. Paper presented at the Annual Digital Government conference, Montreal, Canada.
- Katz, N., Lazer, D., Arrow, H., & Contractor, N. (2004). Network theory and small groups. *Small Group Research*. 35, 307-332.
- Kiesler, S., & Cummings, J. (2002). What do we know about proximity and distance in work groups? In P. Hinds, & Kiesler, S. (Ed.), *Distributed work* (pp. 57-80). Cambridge, MA: MIT Press.
- Kraut, R., Egidio, C., & Galegher, J. (1990). Patterns of contact and communication in scientific research collaboration. In J. Galegher, Kraut, R., & Egidio, C. (Ed.), *Intellectual teamwork: Social and technological bases of cooperative work* (pp. 149-171). Hillsdale, NJ: Lawrence Erlbaum.
- Leydesdorff, L. & Wouters, P. (1999). Between texts and contexts: Advances in theories of citation? (a rejoinder). *Scientometrics* 44, 169–182,
- Liu, Y., Myers, J., Minsker, B., Futrelle, J. (2007), Leveraging Web 2.0 technologies in a Cyberenvironment for Observatory-centric Environmental Research, *OGF-19 Semantic Web 2.0 and Grid Workshop, the 19th Open Grid Forum*, Chapel Hill, North Carolina, USA.
- Malone, T., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys*, 26(1), 87-119.

- Malone, T. W. & Rockart, J. F., (1991). Computers, networks, and the corporation. *Scientific American*, 265, 3, 128-136.
- McNeill, J. R., and McNeill, W. H. (2003). *The human web: A bird's eye view of world history*. New York: Norton.
- McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations: Status distance and the composition of face to face groups. *American Sociological Review*, 52, 370-370.
- McPherson, M., L. Smith-Lovin, L. & Cook, J. (2001). Birds of a feather: Homophily in Social Networks. *Annual Review of Sociology*. 27:415-44.
- Merton, R. K. (1968). *Social Theory and Social Structure*. New York: Free Press
- Metzger, N., & Zare, R. (1999). Interdisciplinary research: From belief to reality. *Science*, 283, 642-643.
- Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. New York: Oxford University Press.
- Monge, P.R., Heiss, B., & Margolin, D.H. (in press, November). The evolution of communication networks in organizational communities. *Communication Theory*.
- O'Leary, M. Orlikowski, W., and Yates, J. (2002). Distributed work over the centuries: Trust and control in the Hudson's Bay Company, 1670-1826. In P. J. Hinds & S. Kiesler (Eds.), *Distributed work* (pp. 27-54). Cambridge, MA: MIT Press.
- Robins, G., Snijders, T., Wang, P., Handcock, M., & Pattison, P. (2007). Recent developments in exponential random graph (p\*) models for social networks. *Social Networks*, 29, 192-215.
- Stokols, D., Fuqua, J., Gress, J., Harvey, R., Phillips, K., Baezconde-Garbanati, L., et al. (2003). Evaluating transdisciplinary science. *Nicotine and Tobacco Research*, 5 (Suppl. 1), S21-39.
- Stokols, D., Harvey, R., Gress, J., Fuqua, J., & Phillips, K. (2005). In vivo studies of transdisciplinary scientific collaboration: Lessons learned and implications for active living research. *American Journal of Preventive Medicine*, 28 (Suppl. 2), 202-213.
- Stringer, M. J., Sales-Pardo, M. & Amaral, L. A. N. (2008). Effectiveness of journal ranking schemes as a tool for locating information. *PLoS ONE* 3, e1683.
- Van de Ven, A., Delbecq, A., & Koenig, R. (1976). Determinants of coordination modes within organizations. *American Sociological Review*, 41, 322-338.
- Wasserman, S. & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Weisband, S. (2002). Maintaining awareness in distributed team collaboration: Implications for leadership and performance. In P. Hinds, & Kiesler, S. (Ed.), *Distributed work* (pp. 311-333). Cambridge, MA: MIT Press.

- Williamson, O. E. (1991). Comparative economic organization: The analysis of discrete structural alternatives. *Administrative Science Quarterly*, 36, 269-296.
- Wuchty, S., Jones, B. & Uzzi, B. (2007a) The increasing dominance of teams in production of knowledge. *Science* 316, 1036 – 1039.
- Wuchty, S., Jones, B., and Uzzi, B. (2007b). Science Commentary: Why Do Team-Authored Papers Get Cited More? *Science*, 317: 1496-1498