This is a preprint of an article published in JASIST (Stvilia, B., Hinnant, C., Schindler, K., Worrall, A., Burnett, G., Burnett, K., Kazmer, M., & Marty, P. (2011). Team diversity and publication patterns in a scientific laboratory. *Journal of American Society for Information Science and Technology*, *62(2)*, 270-283. DOI: 10.1002/asi.21464)

Composition of Scientific Teams and Publication Productivity at a National Science Lab

Besiki Stvilia[1], Charles C. Hinnant[1], Katy Schindler[2], Adam Worrall[1], Gary Burnett[1], Kathleen Burnett[1], Michelle M. Kazmer[1], Paul F. Marty[1]

[1]College of Communication and Information, Florida State University
[2]Department of Sociology, Florida State University
Tallahassee, FL 32306-2100, USA
{bstvilia, chinnant, ks08h, apw06, gburnett, kburnett, mkazmer, pmarty}@fsu.edu

Abstract

The production of scientific knowledge has evolved from a process of inquiry largely based on the activities of individual scientists to one grounded in the collaborative efforts of specialized research teams. This shift brings to light a new question: how the composition of scientific teams impacts their production of knowledge. This study employs data from 1,415 experiments conducted at the National High Magnetic Field Laboratory (NHMFL) between 2005 and 2008 to identify and select a sample of 89 teams, and examine whether team diversity and network characteristics impact productivity. The study examines how the diversity of science teams along several variables impacts overall team productivity. Results indicate several diversity measures associated with network position and team productivity. Teams with mixed institutional associations were more central to the overall network compared to teams composed primarily of the NHMFL's own scientists. Team cohesion was positively related to productivity. The study indicates that high productivity in teams is associated with high disciplinary diversity and low seniority diversity of team membership. Finally, an increase in the share of senior members negatively affects productivity, and teams with members in central structural positions perform better than other teams.

Introduction

The pursuit of scientific knowledge has changed dramatically from the stereotypical sole scholar or investigator often seen in popular media or imagined by the general public. The second half of the twentieth century marked a significant shift in the conduct of science itself and a reorganization of scientific endeavors from the pursuit of knowledge by individuals or small groups to larger projects making use of large numbers of scientists organized into highly specialized research teams. In such environments, team-based collaborations have become the rule rather than the exception, and nowhere is this more true than in large scientific institutions such as national laboratories. In such environments, teams have become an increasingly prominent means of knowledge production. Not only is the size of research teams themselves growing, but so are the knowledge outputs of such collaborations, with multi-authored publications now cited more frequently than single-authored publications (Wuchty, Jones, & Uzzi, 2007). These settings have become important socio-technical environments for examining how the characteristics of scientific teams impact the production of scientific knowledge. As teams form and work together to employ specialized scientific instruments, technologies, and researchers located within the labs, the relationships between team characteristics and knowledge production become important phenomena of inquiry (Lee & Bozeman, 2005; Thorsteinsdóttir, 2000).

Recent years have seen increasing investments and efforts to build an advanced infrastructure for e-science, including high performance computing centers connected through high speed networks (e.g. TeraGrid) to facilitate the wider sharing of expensive science instruments and curated datasets and to enable more effective and efficient scientific collaborations, learning, and professional development (Atkins et al., 2003). For instance, in physics, thousands of scientists from around the world are expected to access and analyze 15 petabytes of data from the Large Hadron Collider (LHC) at CERN (Beagrie, 2006). However, building advanced infrastructure, applications, grids, and collaboratories does not guarantee that scientists will use the technology, or that they will form successful collaborations or teams. Team composition, cultural factors, social structures, and arrangements may either constrain or encourage the adoption and use of technology or data. Similarly, the technology may influence social structures and enable or constrain social interaction, data sharing, and collaboration (Birnholtz & Bietz, 2003; Haythornthwaite, 2002; Orlikowski, 1992; Stvilia, Twidale, Smith, & Gasser, 2008; Wellman et al., 1996), which as a whole may affect team productivity (Haythornthwaite, 2009).

This study addresses these issues by examining the relationships between several measures of team composition and publication productivity of scientific teams at a national science lab, the National High Magnetic Field Laboratory (NHMFL). This article first provides background regarding the institutional environment at the NHMFL, then explores prior findings regarding the relationship between team composition and publication productivity by examining related literature. It then examines the relationships between team demographic and network characteristics (gender diversity, scientific discipline, seniority, and network centrality) and publication productivity, using regression analysis. Finally it discusses the implications of the findings for broader management of scientific teams.

The National High Magnetic Field Laboratory

The NHMFL, the world's largest and most highly powered magnet laboratory, is funded primarily by grants from the U.S. National Science Foundation with additional seed funding and support from the State of Florida. The NHMFL itself is a collaborative venture between three institutions: Florida State University in Tallahassee, Florida; Los Alamos National Laboratory in Los Alamos, New Mexico; and the University of Florida in Gainesville, Florida. Scientific teams apply to use its facilities through a user program, and a review panel including the director of the respective magnet program, NHMFL administrative staff, and subject matter experts evaluates the applications. Selected teams then schedule time to work on the magnet in order to conduct their experimental studies. In-house NHMFL scientists, research staff, and support staff also coordinate with outside teams in order to provide assistance as necessary. Scientific teams do not pay usage fees for work on the magnets themselves; however, they are responsible for a variety of related costs. For example, they must pay any costs associated with their experimental samples and any costs associated with team members traveling to NHMFL facilities.

The only scientific laboratory of its kind in the United States, the NHMFL hosts over 900 scientists per year who use its magnets to run a wide variety of experiments. The Lab is multi-disciplinary, with scientists working on research from a variety of areas in physics, biology, bioengineering, chemistry, geochemistry, biochemistry, and materials science (NHMFL, 2010a). In addition, teams working at NHMFL facilities vary greatly in terms of their characteristics. For example, the composition of scientific teams using the Lab varies greatly in terms of institutional representation as well as scientific disciplines and the seniority of members. The variety of scientific teams at NHMFL makes it a unique environment in which to examine the impact of team composition on the production of scientific knowledge.

Related Work

As scholars and practitioners hypothesized about the presumptive productivity benefits of team-based work, a key area of focus became the question of why some teams are more productive or effective than others. Unfortunately, this seemingly simple question does not have a simple answer, since a number of factors related to team composition and the broader social and technical factors associated with their work environments impact productivity levels. Organization and management researchers have focused attention on the impact of a number of demographic and network characteristics on the productivity of teams performing a number of different core functions such as research and development, strategic management, and production. Similarly, such research has been undertaken in a number of different industrial and institutional contexts such as financial services, health care, and information technology (Joshi and Roh, 2009). The diversity of the prior research regarding team demographic and network characteristics on productivity makes an exhaustive review of the literature beyond the scope of this study. Therefore, the research literature examined here is meant to provide foundational support for the hypothesized relationships we posit between team demographic and network characteristics and productivity within the context of team-based scientific research.

Team Performance and Productivity

Team performance and productivity can be conceptualized and measured in a number of different ways, depending on the context in which the team is evaluated. An examination of product teams in high technology companies relied on two measures, adherence to budgets and schedules and efficacy in developing technical innovations, to evaluate managerial-rated team performance. The same study also employed perceptions of team performance by the team members themselves as a measure of team-based performance (Ancona & Caldwell, 1992). In a study of management team members within a consumer products company, performance was measured as a team's actual profitability relative to its target profitability for the survey year (Bunderson & Sutcliffe, 2002). Still other studies have employed outcome measures to assess the production levels of individual members or teams. For example, studies of open source software projects use the number of completed modification requests as the measure of success (Singh, Tan, & Mookerjee, 2008). Studies of scientific teams often employ measures such as the number of patents or number of publications resulting from the work (Frohlich & Resler, 2001; Marin-Sempere, Garzon-Garcia, & Rey-Rocha, 2008). Conceptualization of performance or productivity is a function not only of the outcome of the task activities but also of the formal and informal goals held by the team or organization.

Institutional Association, Proximity, and Disciplinary Diversity

Collaboration and performance may be associated with both the physical and social proximity between potential collaborators. According to Cronin (2008), the role of place in the choice of collaboration partners has often been undervalued. In a study of his own collaborations and those of Rob Kling, Cronin (2008) found that physical proximity fostered partnerships. Hoegl and Proserpio (2004) studied 145 software development teams to examine the impact of geographic proximity – or place – on team performance, finding that it positively impacted team work quality in five of six facets of work quality: communication, coordination, mutual support, effort, and cohesion; balance of contribution, however, had no significant impact. Cronin (2008) proposed that the role of place in facilitating collaboration may be due to several factors, including "lower transaction costs, ease of coordination, shared organizational culture, swift trust formation, preferential attachment, and sunk cost of investment in co-authors" (p. 1005). Such benefits may outweigh the perceived benefits of more distant collaborations based solely on scientific congruency. Therefore, the benefits of physical proximity may essentially capture the reduced transaction costs associated with institutional proximity.

While place may be in part a proxy for cultural or institutional association, studies directly examining the impact of institutional associations indicate that multi-institutional collaborations experience difficulties with coordination and interaction. A study of principal investigators (PIs) in 62 scientific collaborations (Cummings & Kiesler, 2005) indicates that teams with more multi-university affiliations pose more problems for coordination and have fewer positive research outcomes than teams with fewer such affiliations. In addition, the same study indicates that bringing researchers into closer geographic and physical proximity helps mitigate the negative effects of multi-university collaborations (Cummings & Kiesler, 2005). Therefore, proximity may mitigate cultural and/or task conflict that may flow from having multi-institutional participation in research projects.

Several studies have looked at the relationship between multidisciplinarity and team productivity. Cummings and Kiesler (2005) found that multidisciplinary projects had as many positive research outcomes as those with few disciplines represented. Porac et al. (2004) examined the publication patterns of two multi-institutional scientific project teams, finding that while the collaborations resulted in increased publication for the members of both teams, the productivity increase of the more interdisciplinary team was higher than that of the other.

Gender, Racial, Age, and Tenure Diversity

As with other forms of demographic diversity (such as functional background), gender, age, and racial diversity can function as proxy measures for the shared background and experiences that facilitate team communication and work (March and Simon, 1958; Reagans, Zuckerman, & McEvily, 2004; Zenger and Lawrence, 1989). Gender and racial diversity may be related to increased levels of emotional conflict within teams, because team members do not have similar backgrounds or experiences out of which to develop shared understandings and heightened levels of communication (Pelled, 1993; Pelled, Eisenhardt, & Xin, 1999). However, age diversity is generally negatively associated with emotional conflict within team settings (Pelled et al., 1999), which can impede team process activities such as technical communication. For instance, Zenger and Lawrence (1989) found that within project teams at a U.S. electronics firm age similarity is positively associated with the frequency of technical communication while similarity in group tenure does not impact the frequency of technical communication. In other studies, tenure diversity has had a negative effect on successful team performance (Ancona & Caldwell, 1992; Bunderson & Sutcliffe, 2002). Similarly, similarity in group and organizational tenure – the length of time a person is associated with the group or organization – can enhance the frequency of communication outside of the group itself. Likewise, successful teams tend to have a greater proportion of incumbents, while teams with low tenure diversity are less successful (Guimera, Uzzi, Spiro, & Amaral, 2005). This suggests that shared experiences associated with group or organizational work may facilitate communication that underpins team boundary-spanning activities (Zenger & Lawrence, 1989).

The impact of demographic variables on team performance is not always obvious. For instance, while demographic diversity is often associated with intra-group conflict, Pelled (1996) developed an intervening process theory that categorized demographic diversity variables in terms of visibility (the extent to which the characteristics may be observed by the group) or job-relatedness (the extent to which the variable directly shapes perspectives and skill related to the cognitive tasks) in reference to the group members. Demographic variables such as age, gender, and race are highly visible but have low job-relatedness. Group tenure, however, is highly visible to group members but also has high job-relatedness. Organizational tenure, education, and functional background all have less visibility and high job-relatedness in relation to the group. Both the visibility and the job-relatedness of particular demographic variables influence the level of emotional and task conflict in the team. Similarly, the level of emotional or task conflict can influence turnover and team performance. Diversity among highly visible demographic variables is associated with emotional conflict that leads to negative consequences such as turnover. Consequently, it is proposed that while the potentially positive effects of task conflicts may lead to higher quality and more effective outcomes than result from more job-related variables, emotional conflict may temper these effects (Pelled, 1996).

Seniority

The seniority of team members may also impact intra-group dynamics and productivity. According to Cohen and Zhou (1991), seniority, like other demographic variables such as gender, is an external status characteristic, "one for which social significance is defined prior to and outside of the interaction of interest" (p. 181). Seniority of individual team members is generally conceptualized in terms of the social-hierarchical status of an individual within a broader external group or social network. For example, studies examining the role of scientific teams and productivity sometimes simply measure seniority as the number of years since an individual has achieved a key career milestone such as a doctoral degree (Marin-Sempere, Garzon-Garcia, & Rey-Rocha, 2008). Other studies of research and development teams have measured seniority by ascertaining the amount of time that individuals have worked within a particular team or organization (Cohen & Zhou, 1991). Therefore, the concept of seniority is sometimes measured similarly to the concept of tenure.

As a factor that impacts group interaction and productivity, seniority levels within a team can impact both intra-group interactions and individual status within the team itself. In a study of 224 research and development teams in 29 corporations, Cohen and Zhou (1991) indicate that seniority is positively related to an individual's status within their team but negatively related to the level of interaction the senior team member has with other team members. However, this study also indicated that an individual's status within the broader organization had a greater impact on team status than seniority or other status characteristics such as gender or education level. Seniority is also associated with the level of consolidation and integration of researchers within teams. In a study of scientific teams, Marin-Sempere, Garzon-Garcia, and Rey-Rocha (2008) indicate that senior researchers are more often associated with larger research teams with high levels of consolidation and integration, and that larger teams can often lead to higher levels of research productivity. Conversely, less senior researchers are more often in the process of consolidating or integrating the teams in which they work and have lower levels of productivity.

Network Characteristics

Studies have found relationships between the network characteristics (e.g. centrality, number of incoming and outgoing links and citations) and the performance and reputation of organizations, teams, and individuals (Cronin & Meho, 2006; Haythornthwaite, 2009; Powell, Koput, & Smith-Doerr, 1996; Powell, Koput, Smith-Doerr, & Owen-Smith, 1999). The network properties of a team may also influence an individual's decision to join the team, as well as the overall growth dynamics of the team (Backstrom, Huttenlocher, Kleinberg, & Lan, 2006). Networks may exhibit small-world (Milgram, 1967; Watts & Strogatz, 1998) and scale-free distribution properties (Barabasi & Albert, 1999; Lotka, 1926). Analyzing the network's topology can help identify highly connected components of the network, influential members and groups (e.g. gatekeepers and "hubs"), and the relationships of these structures and properties to information flow, social processes, and team dynamics (Panzarasa, Opsahl, & Carley, 2009; Powell et al., 1999).

Alternatively, social variables and processes may impact the formation of network ties. Powell, White, Koput, and Owen-Smith (2005) tested whether the homophily of nodes along with other characteristics (a node's network degree, popular trends, or diversity of connections) influenced formation of partnership ties between biotechnology companies. Yuan and Gay (2006) investigated whether the similarity of social and demographic characteristics (e.g. race, gender, geographic location) might influence the formation of instrumental or expressive ties in distributed student project teams. Other studies have pointed to the importance of having dissimilarity in a team for successful knowledge creation and exchange of ideas (e.g. Argote & Ophir, 2002).

Team cohesion usually refers to the extent to which the same team members collaborate or interact over time. Singh, Tan, and Mookerjee (2008) defined two types of team cohesion, internal and external, measuring internal cohesion in several ways, including the number of repeat collaborations among team members; they defined external cohesion as a function of network ties among the external contacts of a team. A longitudinal analysis of more than 2,000 open source software projects found a positive interaction between a team's internal cohesion and its productivity. The relationship between external cohesion and team productivity, however, followed a "nominal is best" curve, where project success was the highest for moderate levels of external cohesion (Singh et al., 2008). A study of student project teams, measuring cohesion as a ratio of positive reciprocal communications in the team, also found high performing teams to have high levels of cohesion and lower conflict than other teams (Yang & Tang, 2004). Also, the team longevity associated with prolonged cohesion can moderate the relationship between racial and tenure diversity and emotional conflict within teams (Pelled et al., 1999).

It appears that there is a tradeoff between the similarity and the diversity of team membership, and that teams may need to find a balance to remain successful in providing their functions (e.g. information exchange or knowledge creation). Too much dissimilarity may lead to the balkanization of a team, when the members stop communicating and collaborating or even disrupt each other's efforts (Adamic & Glance, 2005; Stvilia, Twidale, Smith, & Gasser, 2008) while too much similarity may lead to a reduction in the access to resources outside of the team (Reagans, Zuckerman, & McEvily, 2004).

Research Hypotheses

The purpose of the current study is to define an empirically grounded model of the productivity of scientific teams. Guided by an analysis of the literature, it aims to identify and test the relationship between team demographic and network characteristics, and team productivity.

The literature review highlighted several potential relationships between aspects of team diversity and research productivity that are of particular interest to this study. Large scientific organizations such as the NHMFL operate by providing facilities and support staff to large numbers of visiting researchers from a variety of domestic and international research institutions. As research teams form and operate at such large research laboratories, they are often composed of researchers from a number of different research institutions. Since a high level of internal diversity in institutional associations has been associated with increased difficulties in intra-team coordination (Cummings & Kiesler, 2005), they are likely to also suffer from lower levels of research productivity, as diverse teams must learn to overcome such issues to successfully work together. This leads to the following research hypothesis:

H1: Increased diversity in institutional affiliations is associated with a decrease in team research productivity.

While diversity in the institutional associations of team members may inhibit coordination, disciplinary diversity may lead to a broader array of knowledge and intellectual resources available within the team. While such diversity can lead to some initial difficulty in communication due to a wider variety of intellectual backgrounds and value orientations within the team, the increase in intellectual and knowledge resources that can be employed on increasingly interdisciplinary research tasks should facilitate increased team productivity (Cummings & Kiesler, 2005; Porac et al, 2004). This leads to the following hypothesis:

H2: Increased diversity in the scientific disciplines represented in the team is associated with increased productivity.

The literature shows that gender diversity may have mixed effects. Greater gender diversity may lead to higher quality outcomes, especially in cognitive tasks and greater reach and access to resources outside the network. It may also increase the chances of intra-team conflict, as team members may not have similar backgrounds and experiences from which to develop shared understandings and heightened levels of communication (Pelled, 1993; Pelled et al., 1999). This leads to the following hypothesis:

H3: Increase in gender diversity has no or an insignificant effect on team productivity in scientific teams.

Other demographic characteristics also play a role in the productivity of scientific teams. Seniority in research teams can increase the status of the team members but also may decrease their interaction with other team members (Cohen & Zhou, 1991). However, diversity in seniority also is a characteristic of more consolidated and integrated research teams with heightened levels of communication and interaction within the overall team (Marin-Sempere et al., 2008). This leads to the following hypothesis:

H4: Diversity in team seniority is positively related with team research productivity.

Researchers have looked at effects of network position or embeddedness (Granovetter, 1985; Uzzi, 1997) of individual agents and teams on their performance or status in a number of contexts. Studies of college student performance have shown a significant positive relationship between centrality and student performance in class networks (e.g., Baldwin, Bedell, & Johnson, 1997; Yang & Tang, 2004). A study of biotechnology research organizations and firms found connections between centrality of network position and higher levels of innovation (Owen-Smith & Powell, 2004). Studies of open source software teams at SourceForge also found a correlation between network centrality and team productivity (Grewal, Lilien, & Mallapragada, 2006; Singh et al., 2008). This leads to the following hypotheses:

H5: Teams in more central network positions are likely to be more productive.

Methodology

The data for this study consist of a list of experiment teams published in the NHMFL's Annual Reports from 2005 to 2008 (NHMFL, 2010c), and the list of all peer-reviewed publications resulting from NHMFL experiments. The list of publications includes the years 2005 to 2009 and was downloaded from the NHMFL's publication page (NHMFL, 2010b). The researchers extracted 1,415 experiment teams and 2,128 publications from the annual reports and the publication website.

Identifying Teams

A team is usually defined as a small group of agents working together, interdependently to achieve a common goal or complete a shared task (Bell & Kozlowski, 2002; Katzenbach & Smith, 1993, p. 21; Sundstrom, de Meuse, & Futrell, 1990). The tasks and goals of scientific teams, in general, are more uncertain, emergent, and dynamic than those of manufacturing and business teams. Scientists and scientific teams may have short-term goals as well as long-term research objectives and agendas achieved incrementally through individual studies and experiments.

The data from the annual reports included a list of workgroups who developed a research proposal involving one or more experiments, were awarded facility time for those experiments, and then performed the experiments. An examination of experiment teams at the NHMFL indicated that individual scientists, or groups of scientists, may participate on multiple experimental studies. There might be *one to many* relationships between scientific teams and experiment teams. Each team might be involved in many projects, and each project team might be connected with many experiment teams from the annual report list. Hence, the identification of teams from the list of experiment teams involved resolving the problem of duplicate (or almost duplicate) experiment teams – that is, identifying and collapsing duplicate experiment teams into a single team. A good review of the techniques of duplicate or near-duplicate record identification in databases can be found in Elmagarmid, Ipeirotis, & Verykios (2007).

To identify scientific teams, the researchers devised and applied the algorithm from Figure 1 to the set of experiment teams. The procedure described by the algorithm calls itself repeatedly until no duplicate experiment teams are found. To assess the degree of similarity between two experiment teams, the procedure uses a simple "bag of words" approach with a set overlap operator calculated as a Dice coefficient (van Rijsbergen, 1979):

$$c = \frac{2|A \cap B|}{|A| + |B|}$$

where c stands for a Dice coefficient, and A and B are the sets of experiment team member names. The researchers used the overlap operator instead of exact matching to mitigate the effects of instances when an experiment team member set might include a transient member who participated in one experiment but not in the other experiments of the team (e.g., a Lab scientist or technician providing technical support). The overlap threshold at which two teams were considered duplicates was set to 0.75. The threshold value was determined empirically. In particular, both the median and mean values of member set sizes in the experiment teams were equal to four. Setting the overlap threshold at 0.75 ensured that two experiment teams of the size four would be considered as duplicate only if they matched exactly or differed in only one member.

The algorithm identified 594 teams with sizes ranging from 2 to 11 members. The median and mode values (mode value freq. = 240 teams) of the team size distribution were equal to 3.

This is a preprint of an article published in JASIST (Stvilia, B., Hinnant, C., Schindler, K., Worrall, A., Burnett, G., Burnett, K., Kazmer, M., & Marty, P. (2011). Team diversity and publication patterns in a scientific laboratory. *Journal of American Society for Information Science and Technology*, *62(2)*, 270-283. DOI: 10.1002/asi.21464)

```
identifyTeams(experimentTeams)

    initialize experimentTeamsCopy with a copy of experimentTeams
    set duplicateFound with false

    for each experiment team expTeam in experimentTeams
        remove expTeam from experimentTeamsCopy
        initialize simTeams with experiment teams from experimentTeamsCopy with
        the membership overlap with expTeam >= 0.75, and remove those teams from
        experimentTeamsCopy

        if simTeams is not null
            initialize teamCandidate with the intersection of team membership sets in
            simTeams
            store teamCandidate in teams
            set duplicateFound to true
        else
            store simTeams in teams

    if duplicateFound is false
        return teams
    else
        return identifyTeams(teams)
```

Figure 1. The algorithm for identifying *teams* from experiment teams.

Next, the researchers identified sets of publications for each resultant team, matching the member set of each team to the author sets of publications. When deciding which operator to use in matching, the researchers made an assumption that the set of members of a resultant team was a subset (not a superset or an exact match) of the author set of the team's publications. In particular, an assumption was made that there might be instances when a member of a team was not included in the experiment data from the annual report but showed up in the author list of the team's publication. For example, there might be a member who was not involved with the team's experiments but who influenced or contributed to the overall research that lead to the publication. Therefore, the researchers used a subsumption operator to identify a list of publications of a particular team. To be a match, the member set had to be fully subsumed or contained by the author set.

Also, to control for possible effects of the team size on the number of publications, the researchers selected for analysis only the teams of size 3. This reduced the size of the sample from 594 to 89 teams. The reason for setting the size of team to 3 in the sample was twofold: first, three was the median value of team size distribution of the resultant set; and, second, fully connected triples are considered to be building blocks of social networks (Wasserman & Faust, 1994).

This is a preprint of an article published in JASIST (Stvilia, B., Hinnant, C., Schindler, K., Worrall, A., Burnett, G., Burnett, K., Kazmer, M., & Marty, P. (2011). Team diversity and publication patterns in a scientific laboratory. *Journal of American Society for Information Science and Technology*, *62(2)*, 270-283. DOI: 10.1002/asi.21464)

This study did not distinguish between "permanent" (unit) based teams and volunteer teams based solely on shared research interests of its members. The analysis of member affiliations showed that, while some of the teams are comprised of members from different institutions who may collaborate voluntarily because of shared research interests, other teams can be "unit based" and therefore expected to collaborate. For example, the members can be employees of the Lab, researchers from universities, or even the R&D unit of a commercial company. Indeed, the teams comprised only of Lab scientists made up 20% of the sample.

The researchers collected demographic data for each team member, including association with the Laboratory (internal or external), scientific discipline, gender, and seniority. Table 1 displays the coding schemas for the demographic variables along with codes. The coding schemas and most of the values for the variables, with the exception of gender information, were obtained from the annual reports. The rest of the values, as well as gender information, were collected from and determined by triangulating information from multiple sources on the Web, including scientists' homepages, research group or lab webpages, and web directories.

Table 1: The list of composition variable values and codes

Association		Discipline		Gender		Seniority	
Values	**Value Codes**	**Values**	**Value Codes**	**Values**	**Value Codes**	**Values**	**Value Codes**
Internal	0	Biology, Biochemistry, Biophysics	1	Female	0	Undergraduate Student	1
External	1	Chemistry, Geochemistry	2	Male	1	Graduate Student	2
		Condensed Matter Physics	3			Other	3
		Engineering	4			Postdoc	4
		Magnets, Materials, Testing, Instrumentation	5			Technician, programmer	5
						Senior Investigator (NOT a postdoc or student)	6

The demographic variable codes were used to construct team diversity indices, which were calculated as normalized entropies of the member codes of the above variables as follows:

$$H_{normalized} = \left(-\sum_{i=1}^{N} p_i (\ln p_i) \right) / N$$

where p_i is the probability of the i-th code in the team's set and N is the number of codes in the set. Entropy based metrics have been widely used and recommended in the literature to measure diversity with categorical variables (e.g., Ancona & Caldwell, 1992; Cady & Valentine, 1999; Oetzel, 2001; Pelled, Eisenhardt, & Xin, 1999; Teachman, 1980).

The study also constructed a network of scientists based on their co-occurrence in the experiment teams (see Figure 2). To generate the network's graph, the study extracted the complete set of participants of the experiments from the annual reports – 1,514 names. The nodes of the graph represented individual scientists and the edges represented membership of the same experiment team. The graph then was used to develop three network indices for each team:

- Average Betweenness Centrality,
- Average Closeness Centrality, and
- Average Degree Centrality.

Researchers calculated each of these team indices as arithmetic averages of member values of the corresponding network measures (see Tables 1 and 2). Betweenness and Closeness centrality measures are global measures of a node's importance to the network. Betweenness centrality is the fraction of all shortest paths between pairs of other nodes that include the node. A high betweenness centrality score indicates that the node has structural power to serve as a broker and either facilitate or hinder the flow of information and knowledge between the pair of nodes. The closeness centrality score of a node is measured as the number of other nodes in the network divided by the sum of all distances between the nodes and all others. A high closeness centrality position in the network indicates that the node is linked with more people in the network than other nodes, and that it might serve as an information hub or a reference point for others. Alternatively, being close to a higher number of nodes gives a node better and less expensive access to the knowledge and expertise available across the network. Degree centrality, calculated as the number of edges incident to the node, is similar to closeness centrality and is a measure of the node's local importance (Wasserman & Faust, 1994). In addition to the network indices, the study used the number of experiments members worked on together as a metric for measuring team cohesion or the strength of team ties.

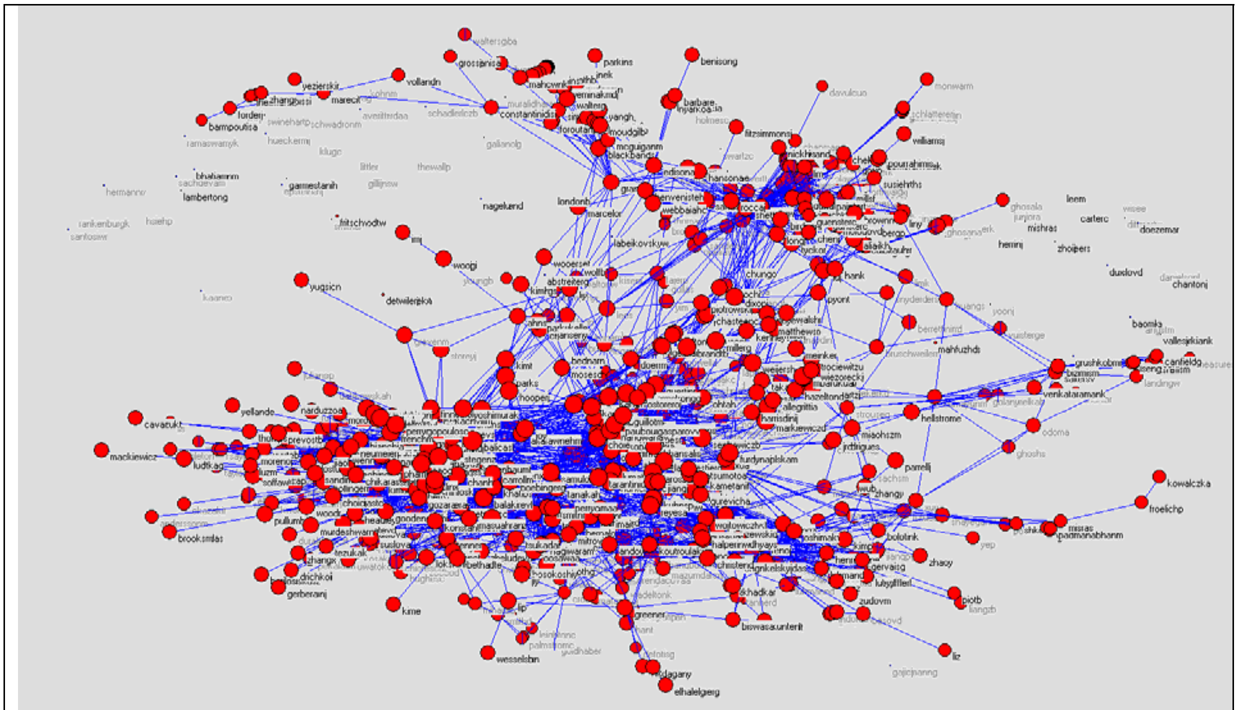

Figure 2. Experiment team member co-occurrence graph; 1,514 nodes; the node sizes reflect member closeness centrality scores.

Table 2: Descriptive statistics of the sample.

	Association diversity	*Disciplinary diversity*	*Gender diversity*	*Seniority diversity*	*Average seniority*
median	0.58	0.58	0	0.58	5
mode	0.58	0.58	0	0.58	4.33
minimum	0	0	0	0	3.33
maximum	0.58	1	0.58	1	6
	Betweenness centrality	***Closeness centrality***	***Degree centrality***	***Cohesion***	***Publication***
median	0.01	0.02	0.25	3	2
mode	NA	NA	NA	1	2
minimum	0	0	0	1	1
maximum	0.04	0.05	0.3	18	16

Before analyzing the data for relationships, researchers tested each team variable/index for normality of distribution using the Wilks-Shapiro Normality test. This test showed that, with the exception of average seniority, the variables were not normally distributed ($p < 0.0001$). This result led the researchers to use nonparametric methods to test for relationships among the variables. In particular, the study used quantile regression (Koenker & Hallock, 2001) to test the hypotheses. In contrast to classical ordinary least squares regression, which estimates the mean of the dependent variable distribution as a function of independent variables, quantile regression allows for the estimation of various proportions (quantiles) of the dependent variable distribution such as median (i.e., 50th) or 75th percentile; therefore, it can provide a more complete view of relationships in a model. Quantile regression is particularly helpful when variables are not normally distributed with heterogeneous variances (Cade & Noon, 2003; Koenker & Hallock, 2001).

Finally, to conduct the statistical analysis the study used Stata software. To preprocess the data, build the experiment membership and co-authorship graphs, extract scientific teams, and calculate team demographic and network indices, the study used Pajek software (http://pajek.imfm.si/doku.php?id=pajek) and Java codes developed by one of the researchers.

Limitations

This study has a number of limitations, some of which will be addressed in future research. To identify the teams and the relationships among team composition and team productivity, it relied exclusively on data extracted from the NHMFL's documents such as annual reports and publication logs. Collecting and analyzing other kinds of empirical data (e.g., observations, interviews, survey) may provide additional insight into to the nature of team relationships. Also, in the analysis, the study controlled for team size and the number of experiments in which members participated; however, it did not control for the length of time teams have been together. The various teams may be at different levels of maturity, which might have an effect on their productivity levels.

In addition, the study assumed that a scientist's demographic status remained unchanged throughout the time period represented by the data sample. However, some of the graduate students might graduate, or post-doctoral fellows might get a faculty or research scientist position. Likewise, a local scientist might leave the Lab and/or be replaced by an outside scientist. Whenever the researchers were aware of such a change in seniority or association during the period of an individual's experiments, we used the earliest seniority or association codes; for example, if a graduate student had become a postdoc during their set of experiments, she or he was coded as a graduate student. In addition, the study followed the coding schema used by the annual reports for the discipline variable. The schema was designed for internal administrative uses and tailored towards the Lab's user community. It may not be as complete as or aligned with some of the "global" disciplinary classification schemas (e.g., Physics and Astronomy Classification Scheme).

The Lab's seniority categorization schema did not differentiate between the academic ranks of scientists (e.g., professor vs. assistant professor). Bibliometric studies of academic productivity suggest that there might be an interaction between the academic rank and the type and level of faculty productivity (Cronin & Overfelt, 1994; Shaw & Vaughan, 2008). The user community of the Lab includes scientists outside of academia and from different countries with different scientific seniority models. Still, replicating this research with a finer granularity seniority schema may provide a more nuanced perspective on the interaction between team seniority and productivity.

The teams comprising the sample were approximations of actual teams. The unavailability of a "gold standard" – the knowledge of the actual team member sets – did not allow the researchers to evaluate the effectiveness (e.g., precision and recall) of the data source (i.e., the annual reports) and the method used to identify those teams. Future work will include the use of surveys and interviews where the researchers will ask the scientists to name their teams, and then compare those teams to the teams identified from the annual reports.

Finally, to evaluate team productivity the study measured the quantity of peer–reviewed publications – an inexpensive measure obtained from the Lab's documents. However, there are other products of scientific teamwork such as patents and non peer-reviewed publications which could be also used in productivity metrics.

Findings

The complete regression model in this study consisted of nine independent variables and one dependent variable. The independent variables included team demographic and structural indices:

- Association diversity,
- Disciplinary diversity,
- Gender diversity,
- Seniority diversity,
- Average seniority,
- Average betweenness centrality,
- Average closeness centrality,
- Average degree centrality, and
- Cohesion.

The dependent variable – number of publications – was used as an indicator of team productivity (see Table 3). Both median and 0.75 quantile regressions (Pseudo R = 0.10; 0.30) of the model found the number of publications to be positively related to team cohesion ($p < 0.005$). Other relationships were not statistically significant (see Table 3).

Table 3: Quantile regression results (* *p<0.05*, ** *p<0.005*)

		Complete Model		Reduced Model	
		0.5 quantile	0.75 quantile	0.5 quantile	0.75 quantile
Variable	Definition	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Association diversity	Normalized entropy of member association codes	0.27(0.66)	-0.34(0.96)	-0.01(0.22)	0.35(1)
Disciplinary diversity	Normalized entropy of member discipline codes	0.61(0.67)	0.52(1.04)	***0.86(0.22)*****	1.25(1.12)
Gender diversity	Normalized entropy of member gender codes	-0.38(0.69)	-0.3(1.03)	-0.31(0.23)	-1.29(1.11)
Seniority diversity	Normalized entropy of member seniority codes	-1.14(1.12)	-1.48(1.43)	***-1.58(0.35)*****	***-4.84(1.75)****
Average seniority	Arithmetic average of member Seniority codes	-0.36(0.49)	-0.29(0.67)	-0.24(0.16)	-1.45(0.77)
Betweenness centrality	Arithmetic average of member betweenness centrality scores in the experiment member co-occurrence network	6.12(49.03)	81.41(79.11)	-3.75(17.18)	-57.11(81.97)
Closeness centrality	Arithmetic average of member closeness centrality scores in the experiment member co-occurrence network	6.83(51.49)	-57.48(85.86)	30.76(17.48)	135.28(83.84)
Degree centrality	Arithmetic average of member degree centrality scores in the experiment member co-occurrence network	0.19(6.77)	8.42(12.14)	-4.06(2.17)	-3.63(13.27)
Cohesion	Number of experiments the members participated in together	***0.37(0.11)*****	***0.73(0.16)*****		
Pseudo R2		0.11	0.30	0.05	0.16

To uncover other less strong predictors that could be obscured by the relationship between cohesion and team productivity, the researchers removed the team cohesion variable from the model and ran quantile regressions again. This reduced the amount of total variance captured by the median regression model from 10% to 5% (see Table 3).

H1 predicted a negative relationship between diversity in team affiliation and team productivity. The statistical analysis found negative but not statistically significant interaction between these two variables. Thus, H1 was not supported.

H2 predicted a positive effect of an increase in disciplinary diversity on team productivity. Results of the analysis revealed positive significant interaction ($p < 0.005$) between the variables, supporting the hypothesis.

H3 predicted no significant interaction between gender diversity and productivity. Although the statistical analysis showed a negative relationship between these variables, the relationship was not statistically significant. Thus, H3 was supported.

H4 predicted a positive interaction between an increase in seniority diversity and team productivity. The statistical analysis, however, found a significant negative relationship ($p < 0.005$) between these variables, rejecting the hypothesis. Also, the analysis suggested that an increase in team seniority level might have a negative effect on team productivity, although the relationship between average seniority and the number of publications fell short of statistical significance ($p = 0.06$).

H5 predicted a positive relationship between high network position and team productivity. The analysis showed mixed results (see Table 3) with the different models, and none of the relationships between the network variables and team productivity was statistically significant. Hence, H5 was not supported.

A 0.75 quantile regression of the reduced model into the number of publications (Pseudo R = 0.15) retained only one significant relationship. Seniority diversity was negatively related to team productivity ($p < 0.05$), with average seniority negatively interacting with team productivity in terms of the number of publications ($p = 0.06$).

The researchers also investigated the relationships among the independent variables by applying Spearman's Rank Correlation analysis to both the complete and reduced models. The analysis showed significant positive correlation between the association diversity index and the average betweenness and average closeness centrality measures. The disciplinary diversity index interacted positively with the average betweenness centrality, while the seniority diversity index positively correlated with the average closeness centrality measure. Also, the average betweenness centrality was the only structural measure showing a significant positive correlation with team productivity in terms of the number of publications. There was no significant interaction between the gender diversity index and any of the network measures (see Table 4).

To better understand the nature of the relationships between member institutional affiliation and the network measures, the researchers developed one more measure, the average association index. This metric was calculated as the arithmetic average of member codes for the association variable (see Table 1). A correlation analysis of the association index with the network indices revealed negative interaction, suggesting that an increase of the share of outsiders in the team might negatively affect the team's centrality in the network. The interaction of the association index with all the network indices was significant at the $p < 0.005$ level (See Table 4).

Table 4: Correlation table of the measures (Spearman's rank correlation; only statistically significant relationships are included; $^{*}p<0.05$, $^{**}p<0.005$).

	Association diversity	Average Association	Disciplinary diversity	Seniority diversity	Number of publications
Betweenness centrality	$+^{*}$	$-^{**}$	$+^{*}$		
Closeness centrality	$+^{*}$	$-^{**}$		$+^{*}$	$+^{*}$
Degree centrality		$-^{**}$			
Cohesion		$+^{**}$			$+^{**}$

Discussion

Different demographic and structural variables may affect scientific collaboration and team performance. Guided by the analysis of the previous literature, this study formulated a set of hypotheses to investigate the relationship among team demographic and network characteristics and team productivity expressed as the number of peer-reviewed publications.

The regression analysis did not provide sufficient support for H1 that teams with diverse affiliations are less productive than homogenous teams. The statistics analyzing the relationship between the association diversity and team productivity were not conclusive. The literature suggests that multi-institutional teams can be less successful than teams with more homogeneous member affiliations due to increased coordination costs (Cummings & Kiesler, 2005). Also, geographic proximity or collocation of team members has often been connected with the higher quality of team communication, higher likelihood of collaboration, and the lower collaboration cost (Cronin, 2008; Hoegl & Proserpio, 2004; Kraut, Egido, & Galegher, 1988). The model of collaboration of multi-institutional teams at the NHMFL is often hybrid. Although a significant amount of experiment design preparation work is done long distance, members of outside teams visit the Laboratory to run experiments, and interact with each other and local scientists face-to-face (personal communication with Kathy Hedick, Chief of Staff of User Programs of the NHMFL, 12 February, 2010). Hence, the strength and nature of their team ties may differ from those of purely virtual teams. In addition, it can be argued that multi-institutional teams might tap into human and technology resources that might not be available at a single institution. Indeed, a recent longitudinal bibliometric study found multi-university collaboration to be the fastest growing authorship trend in scientific communication. The study also found positive interaction between university rank and publication impact (Jones, Wuchty, & Uzzi, 2008). Future work growing out of the current study will explore more context specific structure and tradeoffs of multi-institutional collaborations. In particular, such future work will examine models of work organization used by different types of teams (e.g., virtual, co-located, hybrid), their effectiveness, and different communication and information technologies used in those collaborations using multiple methods: observations, semistructured interviews, and controlled experiments (e.g., Hara, Solomon, Kim, Sonnenwald, 2003; Sonnenwald, Whitton, & Maglaughlin, 2003)..

Results revealed that more interdisciplinary teams may perform better than other teams. The positive interaction between the team disciplinary diversity and team performance was significant, supporting H2. This finding matches findings of some earlier studies (Cummings & Kiesler, 2005; Porac et al., 2004). It is worth noting, however, that the number of publications produced by teams could possibly be affected by differences in the disciplinary patterns of publishing. Future study examining the publication numbers of scientific teams relative to discipline specific baseline rates may shed more light on the nature and underlying structure of the relatively high number of publications by interdisciplinary teams.

The regression analysis did not show a significant relationship between team gender diversity and team productivity, supporting H3. The diversity literature includes mixed findings about the effects of gender diversity on work team performance (Cady & Valentine, 1999; Joshi & Roh, 2009). While some studies have shown that gender diversity increased innovation and quality (Hoffman & Maier, 1961; Rogelberg & Rumery, 1996), others have shown a decrease in productivity, or no direct effect (Cady & Valentine, 1999; Joshi & Roh, 2009; Reagans, Zuckerman, & McEvily, 2004). Also, it has been suggested that the mixed effects of the diversity of demographic variables in general and gender in particular are not unexpected. It has been argued that a change in team diversity can have opposite effects on the strength of a team's internal ties and its external network range. Both of these variables are known to be positively related to team performance. Furthermore, there may be tradeoffs among demographic variables themselves, and, the effects of diversity on team performance may be moderated by the outer context of an organization (Joshi & Roh, 2009; Lawrence, 1997; Reagans, Zuckerman, & McEvily, 2004). For instance, if a field is highly homogenous, a positive effect of an increase in a team's external network range caused by an increase in the team's diversity may be lower due the moderating effect of the field's context, and may not counterbalance the diversity's possible negative effects on the strength of the team's internal ties (Reagans, Zuckerman, & McEvily, 2004). Female scientists comprised only 9% of the total number of scientists in the sample, reflecting the low gender diversity in the field in general (Nelson & Rogers, 2004). Collecting different kind of data through different methods (e.g., observations and interviews) may allow future research to provide further, more nuanced, insight into the interaction between gender diversity and the productivity of scientific teams.

Results showed significant negative interaction between seniority diversity and team productivity, thus rejecting H4. Also, the negative relationship between the average seniority and the number of publications suggested that the presence of students and postdocs in the team may be linked with higher productivity. There seems to be a consensus in the literature that successful teams tend to have a greater proportion of incumbents (e.g., Guimera et al., 2005). There is no agreement, however, on the effects of seniority diversity on team performance. While some studies show a negative relationship between seniority diversity and team success (e.g., Ancona & Caldwell, 1992; Bunderson & Sutcliffe, 2002), other studies have found that teams with higher seniority diversity are more successful (e.g., Guimera et al., 2005). This suggests the possibility of a non-linear relationship between team seniority diversity and team performance. Future research may also explore the structure of the relationship, including effects of organizational culture and formal performance evaluation models on member motivation to produce.

Results did not support H5, which predicted a positive relationship between a team's network position and its productivity. Although the Spearman Rank correlation showed significant positive interaction between the average closeness centrality index and team productivity, the quantile regression analysis did not find significant interaction between the number of publications and any of the network centrality indices. This result is consistent with the results of some past studies, which found positive but non-significant interaction between team closeness centrality and team productivity (Singh et al., 2008).

The study also examined the interaction between team demographic characteristics and network position. The NHMFL is a designated facility of the National Science Foundation with a mission to enable shared use of its unique facilities and instruments by the whole community. Scientists and staff of the Laboratory often serve as enablers and facilitators of experiments run by outside scientists. Hence, it was intuitive and expected to observe the negative interaction between the average association index and the network measures (see Table 4) suggesting a positive relationship between the presence of Lab scientists in a team and team centrality. At the same time, the positive correlation between the association diversity index and the network measures (see Table 4), showed that teams with mixed affiliations were in more central positions in the network than more homogeneous teams (including the teams composed of Lab scientists only). All these suggest that the relationship between institutional association and team centrality might not be linear. The teams with moderate representations of Lab scientists were more central to the network than the teams composed exclusively of local scientists.

Overall, the study linked high productivity in teams to high disciplinary diversity and low seniority diversity of team membership. Also, it was found that an increase in the share of senior members had a negative effect on productivity and that teams with members in central structural positions performed better than other teams.

Conclusions

Teams have become increasingly important in carrying out modern scientific inquiry (Cronin, Shaw, & La Barre, 2004). At the same time, the process of scientific inquiry, in general, is a more uncertain and dynamic process than other kinds of information work. It is usually emergent and less structured. However, even self-organized, self-directed and loosely connected teams may require feedback and/or intervention from outside of the team (e.g., the greater community or administrators) to accomplish their project goals (Douglas & Gardner, 2004; Stvilia et al., 2008). While it is beyond the scope of this work to make definitive policy recommendations regarding the management of scientific teams, this study did define a method and an empirically grounded model of team productivity which could assist in planning effective collaborations and interventions in scientific teams. This study provides evidence to governmental funding agencies, administrators in research laboratories, and the broader science and research policy community regarding the benefits of interdisciplinarity, moderate levels of seniority, and network centrality for the effectiveness of scientific teams. Furthermore, the designers of groupware and collaboration support systems could use the demographic and structural variables of the model and the relationships among those variables in determining desired components and services of a system. The study's findings suggest that including services for analyzing and visualizing the demographic and structural properties of scientists and teams can be useful in planning and facilitating successful teams and collaborations.

The study used the number of peer-reviewed publications as an inexpensive measure of team productivity, which could be easily obtained from the Lab's documents. However, the number of peer-reviewed publications represents only one facet of scientific team productivity. Other facets of productivity could be patents and non peer-reviewed publications. Future research may use more than one metric and source of data for measuring team productivity. Furthermore, future research may also include an analysis of the relationship between team composition and the quality of publications measured by some metric (e.g. the number of citations received).

Both projects and teams are dynamic and may go through different phases from planning and formation to reporting and disbandment (Kerzner, 2003; Tuckman & Jensen, 1977). Each phase may involve different activities and require different kinds of member relationships, skills, knowledge, abilities, and technologies. Also, knowledge of the relationship between team composition characteristics and team performance does not guarantee that members with the desired characteristics will join the team. It has been suggested that the homophily of nodes, along with structural characteristics, may influence the formation of teams and the establishment of specific types of relationship between team members (e.g., Powell et al., 2005; Yuan & Gay, 2006). Future research related to the current study will investigate dynamics of scientific teams, the structure and types of member relationships, and motivations for joining the team. Scientists will be observed and interviewed to collect additional data. All these should help build a more nuanced and comprehensive model of scientific teams.

Acknowledgements

The authors would like to express their gratitude to Larry Dennis, Dean of the College of Communication and Information, for suggesting the NHMFL as a research site and facilitating access to the scientists and staff. Our thanks also to Kathy Hedick for helpful conversations, and to Dr. Gregory Boebinger, director the NHMFL, Bo Flynn, and Pat Dixon for their support of this project. The authors also thank two anonymous reviewers for their helpful comments on an earlier draft of this manuscript. This research was supported in part by the National Science Foundation under Grant

OCI-0942855. The article reflects the findings, and conclusions of the authors, and do not necessarily reflect the views of the National Science Foundation or the NHMFL.

References

1. Adamic, L., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog. In J.Adibi, M. Grobelnik, D. Mladenic, & P. Pantel (Eds.), *Proceedings of the 3rd International Workshop on Link Discovery* (pp. 36-43). New York, NY: ACM.

2. American Physical Society (2010). *Physics and Astronomy Classification Scheme (PACS)*. Retrieved August 16, 2010, from http://publish.aps.org/PACS/.

3. Ancona, D. G., & Caldwell, D. F. (1992). Demography and design: Predictors of new product team performance. *Organization Science*, *3*(3), 321–341.

4. Argote, L., & Ophir, R. (2002). Intraorganizational learning. In J. A. C. Baum, (Ed.), *Companion to Organizations* (pp. 181-207). Malden, MA: Blackwell.

5. Atkins, D. E., Droegemeier, K. K., Feldman, S. I., Garcia-Molina, H., Klein, M. L., Messerschmitt, D. G., Messina, P., Ostriker, J. P., & Wright, M. H. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure.* Retrieved May 25, 2008, from http://www.nsf.gov/od/oci/reports/atkins.pdf

6. Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: Membership, growth, and evolution. In J. F. Elder IV, F. Fogelman-Soulié, P. A. Flach, & M. J. Zaki (Eds.), *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 44-54). New York, NY: ACM.

7. Baldwin, T., Bedell, M., & Johnson, J. (1997). The social fabric of a team-based M.B.A. program: Network effects on student satisfaction and performance. *The Academy of Management Journal, 40*(6), 1369-1397.

8. Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509.

9. Beagrie, N. (2006). Digital curation for science, digital libraries, and individuals. *The International Journal of Digital Curation, 1*(1), 3-16.

10. Birnholtz, J., & Bietz, M. (2003). Data at work: supporting sharing in science and engineering. In M. Pendergast (Ed.), *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work* (pp. 339- 348). New York, NY: ACM.

11. Bunderson, J. S., & Sutcliffe, K. M. (2002). Comparing alternative conceptualizations of functional diversity in management teams: Process and performance effects. *Academy of Management Journal, 45*(5), 875–893.

12. Elmagarmid, Ipeirotis, & Verykios (2007). Duplicate Record Detection: A Survey. *IEEE*

Transactions on Knowledge and Data Engineering, 19(1), 1-16.

13. Cade, B. & Noon, B. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment, 1*, 412-420.

14. Cady, S. &Valentine, J. (1999). Team innovation and perceptions of consideration: What difference does it make? *Small Group Research*, *30*, 730-750.

15. Cohen, B. P. & Zhou, X. (1991). Status processes in work groups. *American Sociological Review, 56*(2), 179-188.

16. Cronin, B. (2008). On the epistemic significance of place. *Journal of the American Society for Information Science and Technology*, *59*(6), 1002-1006.

17. Cronin, B., & Meho, L. (2006). Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, *57*(9), 1275-1278.

18. Cronin, B., & Overfelt, K. (1994). Citation-based auditing of academic performance. *Journal of the American Society for Information Science, 45(2)*, 61−73.

19. Cronin, B., Shaw, D., & La Barre, K. (2004). Visible, less visible, and invisible work: Patterns of collaboration in 20th century chemistry. *Journal of the American Society for Information Science & Technology*, *55*(2), 160-168.

20. Cummings, J. N., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science*, *35*(5), 703-722.

21. Douglas, C., & Gardner, W. (2004). Transition to self-directed work teams: Implications of transition time and self-monitoring for managers' use of influence tactics. *Journal of Organizational Behavior,25(*1), 47-65.
22. Frohlich, C., & Resler, L. (2001). Analysis of publications and citations from a geophysics research institute. *Journal of the American Society for Information Science and Technology*, *52*(9), 701-713.

23. Granovetter, M. (1985). Economic action and social structure - The problem of embeddedness. *American Journal of Sociology, 91*, 481–510.

24. Grewal, R., Lilien, G., & Mallapragada, G. (2006). Location, location, location: How network embeddedness affects project success in open source systems. *Management Science*, 52(7), 1043–1056.

25. Guimera, R., Uzzi, B., Spiro,J. & Amaral, L.A.N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308(5722), 697-702.

26. Hambrick, D. C., Cho, T. S., & Chen, M. (1996). The influence of top management team heterogeneity on firms' competitive moves. *Administrative Science Quarterly*, *41*(4), 659-684.

27. 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.

28. Haythornthwaite, C. (2002). Strong, weak, and latent ties and the impact of new media. *The Information Society*, 18(5), 385-401.

29. Haythornthwaite, C. (2009). Crowds and communities: Light and heavyweight models of peer production. In R. E. Sprague (Ed.), *Proceedings of the 42nd Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society.

30. Hoegl, M., & Proserpio, L. (2004). Team member proximity and teamwork in innovative projects. *Research Policy*, *33*(8), 1153-1165.

31. Hoffman, L. & Maier, N. (1961). Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *Journal of Abnormal and Social Psychology*, *62*, 401-407.

32. Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science*, *322*(5905), 1259-1262.

33. Joshi, A. & Roh, H. (2009). The role of context in work team diversity research: A meta- analytic review. *Academy of Management Journal, 52*, 599-628.

34. Katzenbach, J. & Smith, D. (1993). The wisdom of teams: creating the high-performance organization. Boston, MA: Harvard Business School Press.

35. Kerzner, H. (2003). *Project management: A systems approach to planning, scheduling, and controlling*. (8th ed.). Hoboken, NJ: John Wiley & Sons.

36. Koenker, R., & Hallock, K. F. (2001). Quantile regression. *The Journal of Economic Perspectives*, *15*(4), 143-156.

37. Kraut, R., Egido, C., & Galegher, J. (1988). Patterns of contact and communication in scientific research collaboration. In I. Greif (Ed.), *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work* (pp. 1-12). New York, NY: ACM.

38. Lawrence, B. (1997). The black box of organizational demography. *Organization Science, 8*, 1-22.

39. Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, *35*(5), 673.

40. Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Science*, *16*, 317-324.

41. March, J.G., & Simon, H.A. (1958). *Organizations*. New York, NY: Wiley.

42. Martín-Sempere, M. J., Garzón-García, B., & Rey-Rocha, J. (2008). Team consolidation, social integration and scientists' research performance: An empirical study in the biology and

biomedicine field. *Scientometrics*, *76*(3), 457-482.

43. Milgram, S. (1967). The small world problem. *Psychology Today*, *2*(1), 60-67.

44. National High Magnetic Field Laboratory (2010a). *About the National High Magnetic Field Lab.* Retrieved May 2, 2010, from http://www.magnet.fsu.edu/about/overview.html.

45. National High Magnetic Field Laboratory (2010b).*All peer-reviewed publications by journal name, 2000- 2010*. Retrieved June 16, 2010, from http://www.magnet.fsu.edu/usershub/publications/ReportByJournal.aspx.

46. National High Magnetic Field Laboratory (2010c). *Annual reports*. Retrieved June 16, 2010, from http://www.magnet.fsu.edu/mediacenter/publications/annualreport.aspx.

47. Nelson, D. & Rogers, D. (2004). A national analysis of diversity in science and engineering faculties at research universities. Retrieved August 17, 2010, from http://www.now.org/issues/diverse/diversity_report.pdf

48. Oetzel, J. (2001). Self-Construals, communication processes, and group outcomes in homogeneous and heterogeneous groups. *Small Group Research,* 19-54.

49. Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science*, *3*(3), 398-427.

50. Owen-Smith, J. & Powell, W. (2004). Knowledge networks as channels and conduits: The effects of spillovers in the Boston biotechnology community. *Organization Science, 15*(1), 5-21.

51. Panzarasa, P., Opsahl, T., & Carley, K. M. (2009). Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology*, *60*(5), 911-932.

52. Pelled, L. H. (1993). *Work group diversity and its consequences: The role of substantive and affective conflict*. Doctoral dissertation, Stanford University, Stanford, CA.

53. Pelled, L. H. (1996). Demographic diversity, conflict, and work group outcomes: An intervening process theory. *Organization Science*, *7*(6), 615–631.

54. Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. (1999). Exploring the black box: an analysis of work group diversity, conflict, and performance. *Administrative Science Quarterly*, *44*(1), 1–3.

55. Porac, J. F., Wade, J. B., Fischer, H. M., Brown, J., Kanfer, A., & Bowker, G. (2004). Human capital heterogeneity, collaborative relationships, and publication patterns in a multidisciplinary scientific alliance: a comparative case study of two scientific teams. *Research Policy*, *33*(4), 661–678.

56. Powell, W. W., Koput, K. W., Smith-Doerr, L., & Owen-Smith, J. (1999). Network position and firm performance: Organizational returns to collaboration in the biotechnology industry. *Research in the Sociology of Organizations*, *16*(1), 129-159.

57. Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, *41*(1), 116–145.

58. Powell, W. W., White, D. R., Koput, K. W., & Owen-Smith, J. (2005). Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology, 110*, 1132-1205.

59. Reagans, R., Zuckerman, E., & McEvily, B. (2004). How to make the team: Social networks vs. demography as criteria for designing effective teams. *Administrative Science Quarterly, 49*, 101–133.

60. Rogelberg, S. G., & Rumery, S. M. (1996). Gender diversity, team decision quality, time on task, and interpersonal cohesion. *Small Group Research*, *27*, 79-90.

61. Shaw, D., & Vaughan, L. (2008). Publication and citation patterns among LIS faculty: Profiling a 'typical professor'. *Library & Information Science Research, 30*(1), 47-55.

62. Singh, P., Tan, Y., & Mookerjee, V. (2008). Network effects: The influence of structural social capital on open source project success. *Social Science Research Network: Tomorrow's research today*. Retrieved June 23, 2010, from http://ssrn.com/abstract=1111868.

63. Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. (2008). Information quality work organization in Wikipedia. *Journal of the American Society for Information Science and Technology*, *59*(6), 983-1001.

64. Sonnenwald, D., Whitton, M., & Maglaughlin, K. (2003). Evaluating a scientific collaboratory: Results of a controlled experiment. *ACM Trans. Comput.-Hum. Interact., 10* (2), 150-176

65. Sundstrom, E., de Meuse, K., & Futrell, D. (1990).Work teams. *American Psychologist, 45*, 120-132.

66. Teachman, J. D. (1980). Analysis of population diversity: Measures of qualitative variation. *Sociological Methods & Research*, *8*(3), 341-362.

67. Thorsteinsdóttir, O. H. (2000). External research collaboration in two small science systems. *Scientometrics*, *49*(1), 145-160.

68. Tuckman, B. W., & Jensen, M. A. C. (1977). Stages of small-group development revisited. *Group and Organization Management*, *2*(4), 419.

69. Uzzi, B. (1997). Social structure and competition in interfirm networks: The problem of embeddedness. *Administrative Science Quarterly,* 42, 35–67.

70. Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

71. Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440-442.

72. Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., & Haythornthwaite, C. 1996. Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual Review of Sociology*, 22, 213-238.

73. Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036-1039. doi:10.1126/science.1136099

74. Yang, H. L., & Tang, J. H. (2004). Team structure and team performance in IS development: A social network perspective. *Information and Management*, *41*(3), 335–349.

75. Yuan, Y. C., & Gay, G. (2006). Homophily of network ties and bonding and bridging social capital in computer-mediated distributed teams. *Journal of Computer Mediated Communication*, *11*(4), 1062-1084.

76. van Rijsbergen, C. (1979). *Information Retrieval*. London: Butterworths. Retrieved August 30, 2010, from http://www.dcs.gla.ac.uk/Keith/Preface.html.

77. Zenger, T. R., & Lawrence, B. S. (1989). Organizational demography: The differential effects of age and tenure distributions on technical communication. *The Academy of Management Journal*, *32*(2), 353-376.