Collaboration in an Academic Setting: Does the Network Structure Matter?

Victoria A. Hill
July 14, 2008
CMU-ISR-08-128

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Center for the Computational Analysis of Social and Organizational Systems
CASOS technical report.

This work supported in part by the National Science Foundation under the Graduate Research Fellowship Program, NSF 07-576 and by the Center for Computational Analysis of Social and Organizational Systems (CASOS).

Keywords: Collaboration, Social Network Analysis, Dynamic Network Analysis, Network Structure, Network Metrics

Abstract

What forms of collaboration result in the most benefit to individuals who are in the business of creating new knowledge? I approach this question by examining patterns of collaboration among university faculty members with the objective of determining what types of collaborative relationships are most likely to result in innovative ideas and knowledge creation. By drawing on the toolkits of Social and Dynamic Network Analysis, I measure different structural positions of the network of actors based on this collaborative behavior.

The dataset used in this study contains publication and collaboration data from 1995 to 2006 for each of 61 tenure or research track faculty members in the computer science department of a major U.S. university. Publication data was used as a proxy for knowledge creation. Co-authorship of publications and inter-departmental collaborations on projects, grants and students were used in calculating several network metrics including the E-I Index. These metrics along with relevant control variables are subsequently used in a multivariate regression model to estimate their significance on total publication rates of faculty members. Results indicate that innovation and new knowledge creation are facilitated by new inter-departmental partnerships for a specific cohort of faculty members.

Table of Contents

1. Introduction

The impact that collaborative behavior has on knowledge creation is a topic of great interest of late. However, collaboration can take many forms. The central question is what forms of collaboration result in the most benefit to individuals who are in the business of creating new knowledge? This research endeavor contributes to this topic by examining three specific types of collaborative behavior and assessing their impacts on knowledge creation. Drawing on the toolkits of Social and Dynamic Network Analysis and a dataset of computer science department tenure and research track faculty members of a major U.S. university, different structural positions of the observations are measured based on collaborative behavior. These measures are then subsequently used to estimate their significance on the publication rates of faculty members, which is being used as a proxy for productivity and knowledge creation. The intent is to explore the effects of structural forms of faculty collaboration on knowledge creation, as measured by publication rates. It is argued that network based measures of the structural aspects of collaboration are the most useful.

The theoretical foundation for this approach lies in the precepts of social and dynamic network analysis. Social network analysis focuses on the relationships among social entities, and on the patterns and implications of these relationships, and allows us to examine those patterns in a structural manner (Wasserman & Faust, 1994). Dynamic network analysis extends the boundaries of social network analysis by allowing multi-modal, multi-plex data using the meta-matrix[1] (Carley, 2003).

The dataset used in this study contains publication and collaboration data from 1995-2006 for each of 61 tenure or research track faculty members in the computer science department of a major U.S. university. The approach employed examined the patterns of collaboration among faculty members with the objective of determining what types of collaborative relationships are more likely to result in innovative ideas and knowledge creation. Differentiating the kind of collaborative relationships that facilitate knowledge creation is of utmost importance for policy formulation. Such information would be of interest to department and university administrators, who depend on the reputation of the faculty to attract students and funding.

2. Literature Review

A large body of research has investigated the broad-ranging effects of varying types of relationships on the behavior of both individuals and organizations (Podolny & Page, 1998). Studies have found positive effects of connections on interpersonal power and influence (Brass, 1984), the adoption of innovations (Burkhardt & Brass, 1990; Ibarra, 1993), career opportunities and benefits (Burt, 1992), and creativity (Burt, 2004). Additionally, individuals who have weak-tie relations with many others have been found to gain advantages in employment job searches (Granovetter 1973).

One of the important factors that has been identified in understanding knowledge creation is social capital (Nahapiet & Ghoshal, 1998). They define social capital as the interpersonal relationships of a person, as well as the resources embedded in those relationships. Collaborative relationships can be a means of acquiring social capital. According to Nahapiet & Ghoshal, social capital and knowledge creation will have a positive relationship because social capital

[1] The meta-matrix is a multiple matrix representation of entities and the connections among them. It can examine multiple relationships at once and compute structural measures based on the complete meta-matrix.

directly aids the information exchange process and provides relatively easy access to network resources.

Social capital, however, comes at a cost. Such interpersonal relationships, over time, can eventually limit openness to new information and diverse views (McFadyen & Cannella, 2004). Also, interpersonal relationships take time and effort to create and maintain. There are start-up costs to each relationship (Boissevain, 1974). It is therefore important to determine if certain types of relationships embedded in network structure have greater returns than others.

Collaboration is defined as "working jointly with others or together especially in an intellectual endeavor" (Merriam-Webster, 1999). The importance of academic collaboration is recognized by groups such as the National Science Foundation, which provides funding for a large number of collaborative efforts. This paper defines collaboration as jointly co-authoring a paper, jointly working on an academic grant or project, or jointly advising a graduate student. These collaborations or connections form a social network, and in order to understand their effect, must be looked at from a network perspective.

Network centrality measures are some of the most widely used in network analysis. Perhaps the simplest of centrality measures is *degree centrality* (Wasserman & Faust, 1994). *Eigenvector centrality* (Bonacich, 1972) is a more sophisticated measure of the same concept. Whereas *degree centrality* counts the number of connections an agent has to others in the network, *eigenvector centrality* posits that all connections are not equal to one another. In calculating *eigenvector centrality,* connections to people with more influence in the network count more than connections with people with less influence in the network. *Eigenvector centrality* has been examined in previous research to capture the position or role of an entity in a social network (Podolny, 1994), the structure of inter-organizational collaboration networks (Mizruchi, 1993) and to assess power, prestige and status (Burt, 1992).

Another frequently used centrality measure is *betweenness centrality* (Freeman, 1979). *Betweenness centrality* measures the extent to which an agent facilitates the transmission of information or resources to other agents. *Betweenness centrality* is defined as the number of geodesic paths that pass through the agent. An agent high in *betweenness centrality* has the property of being able to control the flow of information; the gatekeeper or broker. This agent can serve as a liaison between disparate regions of the network. It can signify power, and can be thought of as a measure of the extent to which an agent is positioned to exploit structural holes (Burt, 1992; 2004). An agent who is connected to two unconnected agents has access to resources that both unconnected agents possess, but do not jointly have access to. There can be significant benefits to being in this bridging position. It can reduce redundancy of information; if all of one's ties are to the same group of people, the same information is being recirculated. Granovetter (1973) uses this argument in his discussions about weak ties. He views the network from the point of the tie, or connection, rather than the node, or agent. He argues that weak ties are much more beneficial than strong ties, because when a bridging connection is removed, the agents on either side of the bridge are only reachable via very long paths. Those agents with many weak ties are more important than others due to their positions in the flow of information and resources.

Being in the bridging position allows for the introduction of new and diverse information. Burt (1992) felt that too much cohesion was bad, as it reduced opportunities. He felt that while cohesive groups are able to solve relatively easy problems efficiently, structural holes help to solve more complex problems. Additionally, Burt felt that having connections to different groups increases heterogeneity of ideas and activities, and that breadth of ties should predict innovation and creativity.

The third network variable of interest in this study is *the E-I Index* (Krackhardt & Stern, 1988). In this paper, Krackhardt and Stern introduce the concept of *the E-I Index*. They examine an organization for ties between individuals which have been identified as both external and internal to organizational subunits. In their paper, external links are friendship links between members of different subunits, and internal links are friendships between members of the same subunit. The *E-I Index* is a measure of dominance of external over internal ties, and not simply a measure of external links. The method used by Krackhardt and Stern was a simulation of crisis situations. The results suggest that the structure of relationships in the crisis simulations tested were an important contributor to organizational success. They find that the conditions for successful implementation of major changes include an abundance of ties that cut across formal organizational boundaries. Although the study is focused on organizational change, it can be argued that these same concepts will find much relevance in the computer science field given its fluidic and dynamic nature. Thus I could draw a parallel with the above study by exploring the benefits of collaboration by individual faculty members.

The collaboration examined in the present research is measured in large part from co-authored academic publications. There is a significant body of literature surrounding co-authorship patterns in academia, faculty productivity, and faculty collaborations. However, few studies have been done from a networked perspective. Network studies in academia include an examination of scientific collaboration networks using seven online databases (Newman, 2001). This research showed that the collaboration networks examined exhibit “small world” tendencies, in which randomly chosen pairs of scientists are generally separated by only a short path of intermediate acquaintances. The work further examines the presence of clustering in the networks, and highlights some apparent differences in the patterns of collaboration between the fields studied. It finds that the networks are highly clustered, perhaps indicating the process of scientists introducing their collaborators to one another in the development of scientific communities. However, there is no examination of the relationship of the patterns of collaboration or the rate of publication of the authors. This area presents an opportunity for further research.

Other research has been done examining the number and strength of co-authorship in an academic setting (McFadyen & Cannella, 2004). This research, while not examining how the network structure itself affected returns to co-authorship, found that as relationships increased both in number and strength, returns to knowledge creation diminished. In another study of social capital and knowledge creation, it was found that Burt’s structural holes do not matter (Gonzalez-Brambila, Veloso & Krackhardt, 2006). This surprising result contradicts many earlier studies. Gonzalez-Brambila et. al. found that direct ties, having partners from outside one’s area of knowledge, and being part of a non-dense network had stronger effects on knowledge creation than structural holes.

The present work examines not simply the number and strength of collaborative relationships as in McFadyen and Cannella, but how those relationships compare to one another in the overall network structure, as in the Gonzalez-Brambila et. al. study. It explores collaborations other than co-authorship; collaborations on grants, projects and students and also looks at the differences between collaborations intra-departmentally as compared to inter-departmentally.

It bears stating that previous studies on knowledge creation, collaboration and productivity in the scientific arena have produced results that are often ambiguous and contradictory. This paper attempts to shed some light on this debate by presenting a conceptual scheme from which several hypotheses regarding structural network positioning are presented and empirically verified.

3. Methodology

The dataset employed a cross-sectional dataset consisting of faculty members and the collaborations among them. Much of this network will be characterized by matrices in which the rows and columns represent the faculty members, and the cells represent the collaborations. The core publication data for this study was compiled from a number of different sources. Publication data was collected for each of 61 tenure or research track faculty members in the computer science department, a subdivision of the school of computer science, in a major U.S. university. In examining publications, only peer-reviewed publications were considered with data being collected on total and co-authored publications. Two researchers are considered to have collaborated if they have coauthored a peer-reviewed publication.

Faculty member online c.v.s and websites were the starting point for collecting publication data for 11 years; 1995-2006. That data was then supplemented by a number of online databases including Web of Science, IEEE Xplore, SpringerLink, the ACM Digital Library, CiteSeer and ScienceDirect. The data collection was complicated by several factors. In collecting the data, the sources did not necessarily distinguish between peer-reviewed and otherwise publications. Thus it was necessary to review each publication to make a determination as to whether it was peer-reviewed or not. In addition, some of the online databases do not allow for a full name search; for example, one must input R. Smith, instead of Robert J. Smith. For those databases, it was necessary to physically look at each of the publications in order to determine the correctness of the name. Finally, the sources often returned various versions of the same paper, and these also needed to be examined manually, with duplicates removed. The resulting database is extremely clean, unlike much of the data that has been used in studies of faculty productivity and collaboration using online data sources.

Publication data, and not citation counts, were used as proxies for knowledge creation. While some other research in this area considers citation counts or weighted measures such as the Institute for Scientific Information (ISI) "impact factors," these measures carry their own set of biases. With citation counts, the lag between date of publication and date of citation is uncertain, and this uncertainty limits the use of citation counts only to those publications that have been in print long enough to have had an important impact. This is problematic as publications achieved earlier in time naturally tend to have higher citation counts. Additionally, there is evidence that the impact factors are subject to manipulation and abuse (Monastersky, 2005).

The co-authorship publication data was used to create a binary matrix of *Agent x Agent x Publications* for inter-departmental co-authored publications. Additional collaboration data was collected from the university computer science department head. This additional data tracked faculty member collaborations in the area of grants, projects, and students advised. Each of these three areas of collaboration were used to create a binary matrix; *Agent x Agent x Grants*, *Agent x Agent x Project*, and *Agent x Agent x Students*. All four binary collaboration matrices were then used in dynamic network analysis software, (ORA: Carley & Reminga, 2004; Carley & DeReno, 2006) to create the intra-departmental network metrics of *betweenness centrality* and *eigenvector centrality* for each faculty member. The *E-I Index* was created by identifying the number of unique intra- and inter-departmental co-authors for each faculty member. The *E-I Index*, or *External-Internal Index*, compares the number of internal and external ties among identified groups in the network. For the purpose of this research, the *Internal* metric is the number of unique intra-departmental unique co-authors for that faculty member. Because graduate students may have a big impact on publication rates, the group of graduate students who were in the department during the time period of interest were included in the *Internal* metric. The *External* metric is the number of unique inter-departmental unique co-authors for that faculty member. The *E-I Index* can range from -1, indicating that all of a faculty member's co-authors are intra-departmental, to +1, indicating that all of a faculty member's co-authors are inter-departmental. It is worth noting that this index is a ratio, and is intentionally designed to be independent of network size and density.

Previous research suggests that several other variables may impact on faculty publication rate. Consequently, four additional control variables are added in order to control for these potential effects. The first control variable is *years since PhD*. The second is *gender*, and the third is t*enure vs. research*, and the fourth is *joint* faculty affiliation. Research has been done on life cycle effects on productivity (Levin & Stephan, 1991, Gonzalez-Brambila & Veloso, 2004), with some evidence of decreasing productivity over time. There has also been a significant amount of research describing gender differences in academic productivity, with many suggesting that female faculty members publish less than their male counterparts (Tuner & Mairesse, 2003; Broder, 1993; Long, 1992). Also, faculty members who are on a tenure track have more teaching responsibilities than those who are on a research track. Finally, a variable was added denoting faculty affiliation(s), or *joint*, which identifies those faculty members who have faculty affiliation in more than one department.

These seven metrics; *eigenvector centrality*, *betweenness centrality*, the *E-I Index*, *years since PhD*, *gender*, *tenure vs. research*, and *joint* were then used as predictor variables in a multivariate linear regression to determine their impact on publication rates. Six additional interaction terms and two squared terms were also created.

Based on research previously mentioned, three hypotheses were created regarding possible effects from the three network variables of interest.

Hypothesis 1: Those faculty members who score high in *eigenvector centrality* will have higher publication rates than those who are low in this measure.

Hypothesis 2: Those faculty members who score high in *betweenness centrality* will have higher publication rates than those who are low in this measure.

Hypothesis 3: Those faculty members who score high in the *E-I Index* will have higher publication rates than those who are low in this measure.

4. Data Analysis

The analysis was performed using an OLS multivariate regression model. This model was chosen to determine if any relationship exists between the dependent variable, *newpubs* (publication rate), and the seven explanatory variables *eigenvector centrality*, *betweenness centrality*, the *E-I Index*, *years since PhD*, *gender*, and *tenure vs. research* and *joint.* Prior to the actual regression, the residuals of all of the variables were examined for normalcy. As a result, the three network variables were used in log form; *eigenvector centrality*, *betweenness centrality*, the *E-I Index*. Additional model diagnostics were performed.

Having done with the diagnostics, the regression analysis is performed. Given the functional form:

$$Newpubs = f(\boldsymbol{W}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$$

With an assumed linear relationship specified thus:

$$lnNewpubs_i = \alpha_0 + \boldsymbol{\beta} ln\boldsymbol{W}_i + \boldsymbol{\gamma}\boldsymbol{X}_i + \boldsymbol{\delta}\boldsymbol{Y}_i + \boldsymbol{\lambda}\boldsymbol{Z}_i + \xi_i$$

Where;

$Newpubs_i$ represents total publications over a period of time from 1995-2006, divided by the min of (11, years since Ph.D.);
W, the vector of explanatory variables for the network matrices;
X, the vector of variables that represent observed characteristics of individuals;
Y, the vector of interaction term variables and
Z, the vector of polynomial terms that capture the possible existence of curvilinear relationship
ξ_i is the error term, and the bold face Greek characters represent vectors of coefficients associated with their respective variables.

Three specifications are estimated that progressively expand on the number of the explanatory variables. The base model includes only explanatory variables from both the network metrics and individuals' observed characteristics and is specified thus:

Eqn. 1:

$$lnPub_i = \alpha_0 + \beta_1 lnEigcent_i + \beta_2 lnBetcent_i + \beta_3 lnEiindex_i + \gamma_1 yrsphd_i + \gamma_2 joint_i + \gamma_3 gender_i + \gamma_4 tenres_i + \xi_{1i}$$

The second specification includes the base model and interaction terms. The rationale for this form is motivated by the notion that the explanatory powers of the independent variables may be more pronounced for specific cohorts of the observations. For this regression I include 6 interaction terms. 2 dummy variables – *"tenure research"* and *"gender"* - interacted with each of the 3 network matrixes. For "*tenure research*", a research track faculty member is coded 1 and 0 otherwise with female being the excluded group for the gender dummy.

Eqn. 2:

$$lnPub_i = \alpha_0 + \boldsymbol{\beta} ln\boldsymbol{W}_i + \boldsymbol{\gamma}\boldsymbol{X}_i + \delta_1 Eigcent*tenres_i + \delta_2 Betcent*tenres_i + \delta_3 Eiindex*tenres_i + \delta_4 Eigcent*gender_i + \delta_5 Betcent*gender_i + \delta_6 Eiindex*gender_i + \xi_{2i}$$

The final specification includes all the earlier explanatory variables and the squared polynomial terms. These control variables were included to capture possible curvilinear relationship(s) that may exist between observed characteristics of individuals used in earlier specifications and the dependent variable.

Eqn. 3:

$$lnPub_i = \alpha_0 + \boldsymbol{\beta} ln\boldsymbol{W}_i + \boldsymbol{\gamma X}_i + \boldsymbol{\delta Y}_i + \lambda_1 yrsphdsq_i + \lambda_2 jointsq_i + \xi_{3i}$$

For all these specifications, the error terms are assumed to be normally distributed with zero mean and constant variance.

5. Results and Discussion

Three regression models were submitted that progressively expanded on the number of explanatory variables. Under Specification 1; the base model, findings from the study show that both *betweenness centrality* and the *EI Index* are highly significant predictors of publication rate while *eigenvector centrality* is not – an indication that type of network structural position matters. In this model and in the other two specifications, *years since PhD* remains significant, with a similar negative coefficient. This suggests that there is a negative relationship between publication rate and *years since PhD*.

Specification 2 adds in interaction terms with the three main variables of interest. When I add in the interaction terms to create Specification 2, it is found that *betweenness centrality* is now only marginally significant. Additionally, its coefficient has now doubled in size. *Eigenvector centrality* remains insignificant in both models. This suggests that, at least for this study, attaching oneself to those who are more influential or powerful is not a strategy that will be of benefit in knowledge creation. The *E-I Index* has taken an interesting turn. It is now significant only for those faculty members on the research track, with a higher magnitude, 3.279 as compared to 2.146 in Specification 1. For this cohort of faculty members, the *E-I Index* has a positive effect on publication rate. The coefficient of the *E-I Index* is interpreted such that a 1% increase in the E-I Index has the effect of a 3.279% increase in the publication rate. The reason for this higher magnitude is that Specification 1 looks at aggregates and does mask local realities. The effect for the research track in Specification 1 is diluted. For those faculty members who are on the tenure track, the *E-I Index* is not significant in this model. Interestingly, the *tenure vs. research* variable itself, which is not significant in the base model, is now significant. There is now a small positive effect in moving from the tenure track to the research track.

Specification 3 explores the possible presence of a curvilinear relationship between the dependent variable publication rate and some observed characteristics of individual faculty members. It was observed that none of the squared terms was significant. This shows that a curvilinear relationship does not exist. *Eigenvector centrality* is not significant in Specifications 1, 2, or 3, leading us to reject Hypothesis 1. *Betweenness centrality* maintains its marginal significance in Specification 3, and the coefficient is similar to Specification 2. Therefore, I cannot reject Hypothesis 2 at the 90% confidence level. The *E-I Index* is still significant in Specification 3 for the research faculty, with a similar coefficient to Specification 2. This confirms that irrespective of the stress test that Hypothesis 3 was subjected to, it holds for research track cohort of faculty member.

The results indicate that innovation and new knowledge creation are facilitated by new inter-departmental partnerships for a specific cohort of faculty members; those on the research track and not for those on the tenure track. In none of the models was there a significant return to simply collaborating with those who frequently partner with others. There was no effect found due to gender differences, a result contrary to many previous findings. This could be the result of a very small sample size of women in the dataset; only 7 women out of 61 faculty members.

There has, however, been new research that sheds light on gender differences. Work has been done which finds that it is cultural differences, not gender differences, that matter, and in a culture that emphasizes positive human qualities, both women and men thrive (Blum & Frieze, 2005). It is possible that this department fosters a cultural that allows both men and women to thrive.

There are limitations to this research. It only examines a single department within a school of a university. It is possible that different departments, schools and even universities will have different outcomes. The culture and mechanism of knowledge creation may be different in fields other than computer science. Intuition will lead one to suspect that in a very dynamic field, external links will be more valuable compared to a more static field. Also, there are widely varying criteria for promotion within academia. This research only examines the relationship between collaboration, both intra- and inter-departmental, and publication rate. Other factors, such as service obligations, administrative duties and teaching loads have not been considered.

Areas for further research include investigating academic collaborations on multiple dimensions. It would be interesting to compare regression estimates of the network metrics across departments. A time series regression may provide more robust findings since a panel data set is rich enough to generate not only estimates of levels of the dependent variables but also changes in this variable over time. It also guarantees that time invariant un-observables could be differenced away, thus providing a less noisy regression outcome.

6. Conclusion

It is obvious that collaborations do not come at a zero cost. Thus, the ability to differentiate what kind of collaboration is optimum is of utmost importance and this attribute represents one of the key values added by this research. Findings from this study show conclusively that structural position matters in collaborative networks. It also estimates the contributions to productivity of various forms of collaborations.

The *E-I Index* for research track faculty members was robust across all three models, an invaluable piece of information in formulating policy measures. The insight provided by this finding suggests that any *E-I index* related policy put in place by the department ought to be differentiated since a cohort of its member stands to benefit more from it compared to others. *Betweenness centrality* was marginally significant as well. These findings tell us that it is important to look outwards for innovation and knowledge creation; both intra- and inter-departmentally, and that these qualities are more important than being aligned with another influential agent.

7. References:

Blum, L. & Frieze, C. (2005). As the Culture of Computing Evolves, Similarity can be the Difference. Working paper, Carnegie Mellon University.

Boissevain, J. (1974). *Friends of friends: Networks, manipulators and coalitions.* Oxford: Basil Blackwell.

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

Bonacich, P. (1987). Power and Centrality: A Family of Measures. *American Journal of Sociology, 92*: 1170-1182.

Brass, D. J. (1984). Being in the right place: A structural analysis of individual influence in an organization. *Administrative Science Quarterly*, *29*: 518-539.

Broder, I. E. (1993). Professional achievements and gender differences among academic economists. *Economic Inquiry*, *31*: 116-127

Burkhardt, M. E. & Brass, D. J. (1990). Changing patterns or patterns of change: the effects of change in technology on social network structure power. *Administrative Science Quarterly, 35*: 104–127.

Burt, R. (1992). *Structural Holes*. Cambridge, MA: Harvard University Press.

Burt, R. (2004). Structural holes and good ideas. *American Journal of Sociology, 110*: 349-399.

Carley, K. M. (2003). Dynamic Network Analysis. *Committee on Human Factors, National Research Council*, pp.113-145.

Carley, K. M. & Reminga, J. (2004). ORA: Organizational Risk Analyzer. *CASOS Technical Report*, Carnegie Mellon University, School of Computer Science, ISRI - Institute for Software Research International, CASOS - Center for Computational Analysis of Social and Organizational Systems, CMU-ISRI-04-106

Carley, K. M., & DeReno, M. (2006). ORA: Organizational Risk Analyzer, ORA User's Guide. *CASOS Technical Report*, Carnegie Mellon University, School of Computer Science, ISRI - Institute for Software Research International, CASOS - Center for Computational Analysis of Social and Organizational Systems, CMU-ISRI-06-113.

Freeman, L. C. (1979). Centrality in Social Networks: Conceptual Clarification. *Social Networks, 1*:215-239.

Gonzalez-Brambila, C. & Veloso, F. (2004). The Determinants of Research Productivity: A Study of Mexican Researchers. Working paper.

Gonzalez-Brambila, C., Veloso, F, & Krackhardt, D. (2006). Social Capital and the Creation of Knowledge. Working paper.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*: 1360-1380.

Ibarra, H. (1993). Network centrality, power, and innovation involvement: Determinants of technical and administrative roles. *Academy of Management Journal, 36*: 471-501.

Krackhardt, D. & Stern, R. (1988). Informal Networks and Organizational Crises: An Experimental Simulation. *Social Psychology Quarterly*, *Vol. 51*, No. 2, June.

Levin, S. G. & Stephan, P. E. (1991). Research Productivity Over the Life Cycle: Evidence for Academic Scientists. *American Economic Review 81*, 114-132.

Long, S. (1992). Measures of Sex Difference in Scientific Productivity. *Social Forces*, *71* (1), 159-178.

Merriam-Webster's Collegiate Dictionary (1999). Tenth Edition. Springfield, MA: Merriam-Webster, Incorporated.

McFadyen, M. A. & Cannella, Jr., A. A. (2004). Social Capital and Knowledge Creation: Diminishing Returns of the Number and Strength of Exchange Relationships. *Academy of Management Journal*, *Vol. 47*, No. 5.

Mizruchi, M. S. (1993). Cohesion, equivalence, and similarity of behavior: a theoretical and empirical assessment. *Social Networks*, *15*: 275-307.

Monasterky, R. (2005). The Number That's Devouring Science. *The Chronicle of Higher Education*, *Vol. 52*, Issue 8, October 14.

Nahapiet, J. & Ghoshal, S. **(**1998). Social Capital, Intellectual Capital, and the Organizational Advantage. *The Academy of Management Review, Vol. 23,* No. 2, April.

Newman, M. E. J. (2001). The Structure of Scientific Collaboration Networks. *Proceedings of the National Academy of Sciences of the Unites States of America*, *Vol. 98*, No.2, 404-409.

Podolny, J. M. & Page, K. L. (1998). Network Forms of Organizations. *Annual Reviews Sociology*, *24*:77-103.

Wasserman, S. & Faust, K. (1994). *Social Network Analysis Methods and Applications*, Cambridge University Press.

8. Appendix 1: Tables

Table 1: Summary Statistics

Variable	Mean	Std.Dev.	Min	Max
totalpubs	44.31148	32.78289	3	161
yrssincephd	18	10.08795	2	44
newpubs	4.870795	3.631896	0.272727	18.5
eiindex	0.559603	0.215329	0	1
*eiindex1**	1.559603	0.215329	1	2
gender	0.885246	0.32137	0	1
tenres	0.213115	0.412907	0	1
Joint	1.47541	0.648243	1	4
betcentu	0.01738	0.027504	0	0.1298
*betcentu1**	1.01738	0.027504	1	1.1298
eigcentu	0.014553	0.022543	0	0.075
*eigcentu1**	1.014552	0.022543	1	1.075
l_newpubs	1.297584	0.838826	-1.29928	2.917771
l_eiindex1	0.434559	0.143843	0	0.693147
l_eigcentu1	0.015994	0.02306	0	0.078441
l_betcentu1	0.017478	0.027137	0	0.13173
eiggender	0.014204	0.022536	0	0.078441
betgender	0.015019	0.026409	0	0.13173
eiindexgender	0.38557	0.198227	0	0.693147
eigtenres	0.00034	0.001011	0	0.004191
bettenres	0.001291	0.004488	0	0.020978
eiindextenres	0.083966	0.182505	0	0.693147
yspsquared	422.4754	439.3255	1	1936
jointsqd	2.590164	2.571491	1	16

*a transformation of these variables was performed by adding 1 before the log version was taken

Table 2: Correlation Table

	l_newpubs	*l_eigcentu1*	*l_betcentu1*	*l_eiindex1*	*yrs~phd*	*gender*	*tenres*	*joint*
l_newpubs	1							
l_eigcentu1	0.0887	1						
l_betcentu1	0.3416	0.3371	1					
l_eiindex1	0.4435	-0.3113	-0.0129	1				
yrssincephd	-0.3785	-0.0334	0.0494	-0.26	1			
gender	-0.0555	0.0062	-0.0529	0.0193	0.0874	1		
tenres	-0.2939	-0.3276	-0.2209	-0.148	0.068	0.1874	1	
Joint	0.0637	-0.2128	0.0728	0.2972	0.209	0.0262	0.0735	1

Table 3: OLS Regression results

Explanatory Variables	**Specification 1 (Base Model)**	**Specification 2 (Base Model + Interaction terms)**	**Specification 3 (Base Model + Interaction terms + Squared terms)**
Eigenvector Centrality	1.403	-2.288	-0.843
	(4.689)	(11.659)	(12.026)
Betweenness Centrality	9.827***	18.023*	18.888*
	(3.501)	(9.860)	(10.011)
Eiindex	2.146***	0.534	0.264
	(0.732)	(2.613)	(2.656)
Years Since Ph.D.	-0.024**	-0.022**	-0.027
	(0.010)	(0.009)	(0.031)
Joint (Faculty Affiliation)	-0.014	0.055	0.560
	(0.151)	(0.141)	(0.540)
Gender	0.014	-0.084	-0.202
	(0.281)	(1.267)	(1.288)
Tenure/Research track	-0.282	-2.267***	-2.225***
	(0.241)	(0.674)	(0.683)
Interaction Terms			
Eigenvector x Gender		2.481	1.455
		(12.481)	(12.771)
Betweenness x Gender		-10.408	-10.932
		(10.521)	(10.703)
Eiindex x Gender		0.692	1.060
		(2.735)	(2.800)
Eigenvector x Tenure/Research Track		237.272	225.797
		(154.617)	(159.01)
Betweenness x Tenure/Research Track		35.432	38.872
		(31.255)	(31.912)
Eiindex x Tenure/Research Track		3.279**	3.170**
		(1.329)	(1.350)
Squared Terms			
Joint (Faculty Affiliation) squared			-0.013
			(0.138)
Years since Ph.D. squared			0.000
			(0.000)
Constant	0.668	1.107	0.792
# of Observations	61	61	61
R- Square	0.4233	0.5732	0.5822

Standard errors of coefficients are in parentheses. ***p<0.01, **p<0.05, *p<0.1

9. Appendix 2: Network Variable Calculations

Eigenvector centrality is calculated as follows: denoting the centrality of vertex i by x_i, I reflect this by making x_i proportional to the average of the centralities of i's network neighbors:

$$X_i = \lambda^{-1} \sum_{j=1} A_{ij} X_j$$

The mathematical expression could be stated in matrix form thus:

$$\boldsymbol{\lambda x = A.x}$$

Where **A** is the adjacency matrix of graph G, λ is the largest *eigenvalue* and **x** is the corresponding normalized eigenvector.

Betweenness centrality is calculated as follows: for a graph G, with n vertices, the *betweenness centrality* $C_B(i)$ for vertex i is:

$$C_B(i) = \sum_{u \neq i \neq w \in I} \frac{\sigma_{uw}(i)}{\sigma_{uw}}$$

Where σ_{uw} is the number of geodesics between u and w, and $\sigma_{uw}(i)$ is the number of geodesics between u and w that passes i.

The ***E-I Index*** is calculated as follows:

$$\frac{E - I}{E + I}$$