STUDIES ASSOCIATION

INDUSTRY STUDIES ASSOCATION
WORKING PAPER SERIES

Social Capital and the Creation of Knowledge

By

Claudia N. Gonzalez-Brambila
Instituto Tecnológico Autónomo de México
Mexico City, D.F., 01000, Mexico
cgonzalez@itam.mx

and

Francisco M. Veloso
International Motor Vehicle Program
Cambridge, MA 02139
and
Carnegie Mellon University
Pittsburgh, PA 15213
fveloso@cmu.edu

and

David Krackhardt
Carnegie Mellon University
Pittsburgh, PA 15213
krack@andrew.cmu.edu

2008
Industry Studies Association
Working Papers

WP-2008-18
http://isapapers.pitt.edu/

1. Introduction

The increasing importance of knowledge as the pillar of contemporary economies has been largely recognized over the last decades. The production, dissemination and use of knowledge have become a crucial factor in enhancing economic growth and social welfare. Within this context, science and technology (S&T) are one of the critical means of knowledge creation. Universities and research institutes in particular, play a vital role in the creating and transmitting of scientific knowledge. However, we know surprisingly little about how such crucial knowledge is created (McFadyen and Cannella 2004). The “ontological” dimension of knowledge creation assumes that, although ideas are developed by individuals, the interaction between individuals typically plays a critical role in articulating and amplifying that knowledge (Nonaka 1994). Thus, it is not surprising to find a growing attention to the role of collaborative effort in the process of scientific knowledge generation (Stephan and Levin 1997). Most existing research in this area has looked at issues such as size of scientific teams, institutional collaboration and the geographic dispersion of team members (Adams et al 2005; Stephan and Levin 1997; Wang, et al 2005; Melin 1999; Melin 1996; Barnett et al 1988; Katz 1994). However, these studies have not analyzed relational and structural dimension of the collaborative networks involved in scientific knowledge generation.

When analyzing knowledge creation, one of the factors that researchers have noted as critical is social capital[1] (Nahapiet and Ghoshal 1998). The positive relation between social capital and knowledge creation is typically explained by combination and exchange processes, where the combination takes the Schumpeterian view of the foundation for economic development, and exchange refers to social interaction and coactivity that creates new knowledge (Nahapiet and Ghoshal 1998). This paper examines precisely the relationship between social capital and knowledge creation in research, mostly in the context of universities. For this purpose, knowledge creation is measured by research papers in internationally

[1] Social capital is a function of social structure producing advantage for individual or groups (Coleman 1988).

peer-reviewed publications and social capital is measured through the pattern of connection between actors, where a connection between two researchers is established through co-authorship.

To the best of our knowledge, this is the first study to bring together most of the critical aspects of social capital: direct ties, strengths of direct ties, density, structural holes, centrality, and external-internal index in terms of fields of knowledge. The theories behind these dimensions have been sometimes contradictory (Burt 1992; Burt 2001; Ahuja 2000), and empirical studies have not examined the relative contribution of each dimension, especially in an environment like academic research. Moreover, this study explores the importance of relationships at the individual level, rather than at the organizational level, and considers an entire network. Most network studies have found it difficult to adequately specify the boundaries of the network (Gulati 1995). In this study, all authors that have published with a Mexican author in the area of Exact Sciences are considered; thus, the analysis is based on a very complete network. This is also the first research that looks at scientific collaboration outside the developed world.

Another crucial characteristic of this research is that it uses recent and extensive panel data that allows us to control for individual unobserved heterogeneity. This is important because it is reasonable to assume that individual characteristics that are unobserved by the econometrician, for example the intrinsic quality of individual as measured by intellectual ability, might be correlated with particular network dimensions. In addition, different controls for time varying individual characteristics that could be confounded with the critical network variables are also included. Most existing studies in the network literature have not been able to carefully analyze within differences (Burt 2001) and, as it turns out, controlling for individual specific heterogeneity has a major impact when looking at the effects of network in the performance of individuals.

Two important results arise from this research. First, when controlling for other network variables and individual heterogeneity, the effects of the structural holes variable is negative or disappears. This result stands in contrast to the established idea that structural holes is the most important variable to represent social capital and, therefore, is seen as contributing to superior performance (Burt 1992 and 2001, and Krackhardt 1999, Reagans and Zuckerman 2001). Second, the results show that with this strong set of controls, what matters in social capital is having many direct ties, being in a central position, having partners from different areas of knowledge, and being part of a non dense network. These results have important implications for the configuration of optimal network. Not only we showed, as other authors have found (Gulati 1998, Rowley et al 2000), that both relational and structural embeddedness influence individual performance, but also we present evidence on how individuals could be strategically embedded in the academic web.

The structure of the paper is as follows. The next section provides some theoretical background of social capital and knowledge creation. The subsequent section presents the data used as empirical analysis, and describes the variables as well as the models specifications of the study. Section four presents the results, while the final sections have a discussion and then conclude.

2. Theoretical Background

"The social capital metaphor is that people who do better are somehow better connected" (Burt 2001). However, there is no compelling evidence of what better connections mean, and what kind of network structure enhances the creation of new knowledge. In the social network analysis literature, two opposing versions of the theory of network structure coexist. According to one view (Coleman 1988) actors in embedded networks have superior achievements because members obtain more coordination, they trust each other and develop better communication skills. An alternative view (Burt 1992) suggests that actors who are connected to others who are not connected to each other, that is, open social structures with many structural holes, can take advantages of the "bridges" to connect with new members in other clusters, and get access to new information.

A different set of views on this issue confers great value to the identification of the "most important" actors (Freeman 1982; Wasserman and Faust 1994), i.e. those in a strategic location with many close[2] relationships. The idea is that these actors have advantages because they can get access and transmit new

[2] We refer as close relationships based on the distance or number of paths between the focal actor and his alters.

information sooner than actors in the periphery. Finally, other scholars (Granovetter 1973; McFadyen and Cannella 2004) confer importance to the relational dimension of networks, typically looking at the number and strength of direct ties, often regardless of any embeddedness or centrality issues of the network structure.

Much of the empirical research considering performance-related outcomes has focused on the structure of networks (Burt 1992), and less attention has been paid to the effects of the relational dimension of networks (Cross and Cummings 2004). Yet it is not clear that the analysis of network structure alone captures the effects of the relational dimension for the creation of new knowledge (Cross and Cummings 2004). For example, in a very recent study, McFadyen and Cannella (2004) use a sample of publications of 173 biomedical scientists from 2 universities to test the relationship between social capital and knowledge creation. They find that superior knowledge creation is associated to an early increase in the number and the strength of direct relations, though with diminishing returns, leading to an inverted U relation between these variables and performance. Although the evidence offered by this study provides important insights, the study has some limitations. As acknowledged by the authors, the sample is small and only considers biomedical scientists, so it may not be representative. Besides, the study is limited to direct and strength of direct ties, but in turn ignores embeddedness or centrality concerns.

We argue that to better understand the relationship between the creation of new knowledge and social capital, it is necessary to analyze structural as well as relational dimensions of social capital (Nahapiet and Ghoshal 1998). In particular, bearing in mind a notion that access to new information is the most important direct benefit of social capital (Inkpen and Tsang 2005) we consider most critical variables that can potentially influence the access to new information along both dimensions.

The first relational dimension we consider, are the number of direct relationships. These are expected to stimulate combination and exchange of resources within the relationships (Nahapiet and Ghoshal 1999) and provide researchers with access not only to new knowledge but also to new experiences. Thus, it can be expected that an increase in the number of direct ties will increase the amount of knowledge, ideas, resources people have access to and thus enhance their ability to address complex problems (McFayden and Canella 2004; Reagans and McEvily 2003; Ahuja 2000). Hence, we expect,

Hypothesis 1. Researchers with a larger the number of direct ties will publish more.

A second dimension of relational social capital is the strength of the relationships. Research looking at these issues has ambiguous perspectives. For example, strong and weak ties[3] are argued to both provide benefits, albeit of different nature (Rowley et al 2000). On one hand, weak ties are argued to bring novel information, thus contributing to the creative process (Granovetter 1973; Rowley et al 2000). On the other hand, strong ties promote trust, reciprocity, long-term relationships and the transfer of high quality information and tacit knowledge (Reagans and McEvily 2003, Rowley et al 2000; Gulati 1995; Larson 1992; Inkpen and Tsang 2005). Thus, it is not obvious the strength of ties that is preferred in an environment like academia. It seems that the benefits that is argued to have either weak and strong ties could be obtained, so the purpose of our second hypothesis is to test which kind of tie, weak or strong, contribute to increase the productivity of researchers.

Hypothesis 2. Researchers with strong relationships will publish less.

As noted above, in addition to relationship dimensions, we also look at network structure. Coleman's (1988) main argument is based on the notion that members in dense networks can secure the benefits of getting access to information because they develop trust and shared norms of behavior that mitigate potential opportunistic behavior. However, as mentioned by Burt (1992), even if such knowledge sharing occurs, after some time this information will become redundant. Thus, he rather argues that actors

[3] According to Granovetter (1973) " the strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie".

embedded in sparsely connected networks (rich in structural holes) can take advantage of knowledge brokerage opportunities created to construct an efficient, information-rich network, where the redundancy between partners is minimized (Burt 1992). Efforts have been made (Burt 1998; Burt 2000; Rowley et al 2000; Nicolaou and Birley 2003) to show that these two forms of network structure might not necessarily contradictory, but that both offer advantages depending on the conditions and populations. Thus in accordance to this perspective, we conjecture competing hypotheses based on Coleman's notion:,

Hypothesis 3. Researchers embedded in dense networks will publish more

as well as in Burt's notion:

Hypothesis 4. Researchers embedded in networks rich in structural holes will publish more.

It has been found (Hanneman 2001; Cross and Cummings 2004) that position in the network might affect the opportunities and constrains of an actor. This might be particularly relevant in context of academia because of the highly skewed nature of publications, citations and overall academic prestige (Lotka 1926) around a few select number of researchers. Thus, it is particularly important what Wasserman and Faust (1994) call prestige measure of centrality, in which the centralities are recursively related to the positions to which they are connected. As suggested by Reagans and McEvily (2003), we expect this kind of centrality to increase people's perspective and enhance their ability to tackle complex ideas, thus contributing to increase researchers' productivity.

Hypothesis 5. Scientists with high level of exposure to other central scientists will publish more.

Finally we also explore knowledge clustering, looking in particular at the degree of collaboration among scientists within the same discipline versus collaboration of scientists from different disciplines. Because researchers are addressing more complex research questions, we have seeing an increasing division of labor in the profession (Arora and Gambardella 1990). But this makes the participation of scientists from different disciplines necessary to have access to complementary assets and skills. Thus, we expected that

Hypothesis 6. Scientists that collaborate with researchers from different fields of knowledge will publish more.

3. Method

Our main proposition is that to better understand the impact of social capital in the creation of new knowledge, it is necessary to analyze the effects of all relevant network variables. The components of social capital are many and varied, and closely related to each other. Therefore, the isolated effect of one variable could be very different when the other components are jointly analyzed.

3.1 Data

We analyze the hypothesis described above using a database of publications and citations for all scientific papers that have at least one author from Mexico, published between 1981 and 2002, included in the Science and Social Sciences Citation Indexes produced by the Institute of Scientific Information (ISI) (ISI 2003). Figure 1 shows the evolution of Mexican publications indexed in ISI. As can be seen, there has been an important growth in the number of publications, mainly during the 1990's.

The data on publications record:

- Number of citations
- Date of publication
- Name of authors
- Address information
- Number of coauthors
- Field of knowledge

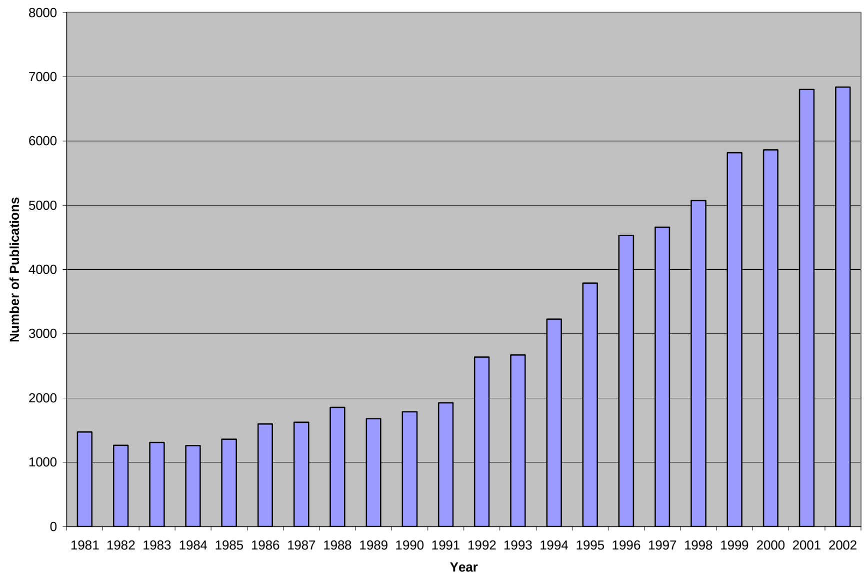

Figure 1. Evolution of Mexican Publications, 1981-2002

In addition, we had access to information on 14,328 researchers, in all fields of knowledge, who have been part of the Mexican National System of Researchers[4] (SNI) from 1991 to 2002.

The data on SNI researchers include:

- Name of researcher
- Gender
- Age
- Country where PhD was earned
- Area of knowledge
- Number of papers published in ISI[5]

[4] The Mexican National System of Researchers was created in 1984 to enhance the quality and productivity of researchers in Mexico. It gives pecuniary compensation, as a complement of salary, to the most productive researchers.

[5] The publications were obtained by matching the database of the researchers in SNI with Mexican articles from the ISI database from 1981-2002 (ISI 2003).

Given that we lack personal information about researchers outside the SNI system and recognizing the great differences in productivity among areas of knowledge (Gonzalez-Brambila & Veloso 2007), the analysis in this paper is restricted to a sample of 1704 researchers in Exact Sciences[6] that have been part of the system at least one year from 1991 to 2002. It is important to stress that all authors in Mexican publications were considered to establish the network variables used in the estimation. Nevertheless, the focal analysis and conclusions are associated to the networks in which SNI researchers in Exact Sciences participate. In 2002, 30% of the Mexican Researchers were part of SNI (Conacyt 2003). Yet, these published almost 90% of the Mexican publications in ISI.

Since the main purpose of this paper is the analysis of social capital in the creation of new knowledge, we dropped from the calculation of network variables analysis all publications with more than 8 authors.. We believe that publications with more than 8 authors reflect other type of collaboration effort and not necessarily the "actual" social capital of researchers. Most papers with many authors are the results of large projects that are conformed from separate contributions that are made in smaller groups.

Co-authorship in ISI publications is used to measure relationships. Melin and Persson (1996) find that a significant proportion of scientific collaboration leads to co-authored papers, and publications in ISI are the most common measure of scientific productivity (Levin and Stephan 1991; Stephan and Levin 1997; Turner and Mairesse 2003; Gonzalez-Brambila and Veloso 2007). Yet, an important limitation of using this measure is that it does not directly measure actual contact between people. On one hand, not all collaborations end in a publication in ISI; and on the other, there are other outputs that can be the result of collaboration among researchers and those are not reflected in our dataset. There are other forms of collaboration that a bibliometric study is not able to reveal. However, by using this measure we avoid the subjective bias of interviews.

[6] This is the official classification in Conacyt, the National Council for Science and Technology in Mexico. It falls within natural Sciences and includes Physics, Mathematics, Astronomy, Geology, Oceanography, Geophysics, and Material Science.

3.2. Variables

In this paper we measure knowledge creation by using the straight number of publications because it is a documented new knowledge and is an indicator of the advances in a field of research (Stephan and Levin 1991). Moreover, researchers have strong incentives to create and disseminate new knowledge through publications, since they are rewarded by doing that (Stephan 1996). Yet, since there might be variations in the quality and impact of the published papers (Lindsey 1989), we also consider the number of cites that each publication received in the subsequent 4 years.

Since most people who have written a paper together will know each other quite well, we considered that two researchers are connected if they have a coauthored paper. To study the network of scientist two-mode matrices were built, including all authors (within and outside SNI) that have published with SNI researchers in Exact Sciences in a given period. To get a one-mode matrix that relates all authors with their publications, the cross-products (co-occurrence) method was used (Bogartti et al 2002; Wasserman and Faust 1994).

The network variables used were:

- Direct ties is the number of unique coauthors. This is the sum of the rows in the dichotomized adjacent matrix with no diagonal.
- Strength of direct ties is the frequency with which an author publishes with their partners. This is the sum of the rows in the adjacent matrix divided by the direct ties.
- Structural holes is the separation of different actors who are not connected. This variable is obtained by subtracting 1 – Constraint. The constraint is obtained through Burt's formula (1992). In essence, constraint is a measure of the extent to which ego is invested in people who are invested in other of ego's alters.
- Centrality is measured by using the normalized eigenvector proposed by Bonacich (1972). This measure was chosen over other measures of centrality because the formula recursively implies that the centrality of an actor is proportional to the centralities of the nodes he is connected to.

Thus, a high value is given to an actor who is connected to many actors who are themselves also well-connected (Bogartti 1995).

- Density, is the number of ties divided by the number of pairs, times 100, where pairs is the total number of pairs of alters in the ego network - i.e., potential ties (Bogartti et al 2002)
- The external-internal (E-I) index is the number of ties external to the groups minus the number of ties that are internal to the group divided by the total number of ties (Krackhardt and Stern 1988). To this purpose, an attribute vector that classifies researchers in fields of knowledge[7] was built. Given that we lack information on the discipline of researchers who have not been part of SNI and because a given researcher could have publications in more than one field of knowledge, it was assumed that the first publication in our records determined the field of knowledge of researchers.

All these network variables were calculated using UCINET 6 (Bogartti et al 2002).

3.3. Models

To assess the effects of network structure in the creation of knowledge, it is assumed that the function determining publishing proficiency P_{it} is given by:

$P_{it} = F(X_{it-1}, c_i, u_{it-1})$, i identifies researchers and t period.

X_{it-1}: Variables that vary across time and across researchers:

Number of direct ties, strength of direct ties, structural holes, density, centrality, external-internal index, reputation (see below for description).

c_i: is the individual unobserved effect which is stable across time but not across researchers

u_{it-1}; is the error term

[7] The classification was done by considering 11 different fields of knowledge that matches the 105 ISI categories into subject groups. This mapping was based on an analysis of journal usage by researchers working in different subject departments in UK universities (Adams 1998).

We use the negative binomial fixed effects model proposed by Hausman, Hall and Griliches (1984) because of the panel nature of the data. The fixed effects model allows for both the possibility of permanent unobserved individual effect as well as the possibility that some unobserved effects may be correlated with publications and other explanatory variables. To assure consistency, a Hausman test (Hausman 1978) was run to check for the possibility of a random effects panel structure. Given the significance of the P-value, we restrict our analysis to the use of fixed effects. The Negative Binomial distribution was chosen over the Poisson because the latter imposes a constant variance. This is not true for the data used in our study where the variance of productivity far exceeds the mean. One drawback of the Negative Binomial distribution is that the conclusions may be less precise since the estimated standard errors tend to be larger than in the alternative Poisson model.

Two different measures of research output were created. The first one (pubs) measures the straight publications counts occurring over a two-year period. The second one (cites) adjust these publications for quality considering the number of cites that publications have received in the subsequent 4 years (Gonzalez-Brambila & Veloso 2007). We decided to consider publication output over 2 years because publication is an uneven event and many researchers do not publish every year. By using this measure we avoid losing observations from having zeros in the outcomes.

The network variables for researcher i at time t-1 were obtained from the adjacent matrixes[8] considering information on the publications in the previous 3 years. The choice of using publications in the previous 3 years meant to balance, research projects' time frame with having enough observations over time. Yet, for robustness we also considered an equivalent analysis to the one reported here but using 4 and 5 year windows. No significant differences were found in the coefficients of the critical network variables that we are interested in analyzing.

[8] The "adjacency" matrix is a matrix composed of as many rows and columns as there are researchers, and where the elements represent the ties between actors, this is the number of joint publications.

Thus, nine periods were obtained as is showed in the following table:

Period	**T**	**t-1**
9	2002/2001	2000/1999/1998
8	2000/1999	1998/1997/1996
7	1998/1997	1996/1995/1994
6	1996/1995	1994/1993/1992
5	1994/1993	1992/1991/1990
4	1992/1991	1990/1989/1988
3	1990/1989	1988/1987/1986
2	1988/1987	1986/1985/1984
1	1986/1985	1984/1983/1982

In an effort to isolate the effects of network structures in productivity a number of control variables was included. Considering the large growth in the number of publications (Figure 1) time dummies were included to capture time trends. However, it is important to note that including time dummies prevents us from adding researcher age or time since PhD in the regressions as a control due to collinearity problems. In addition, we also include a control for changes in researcher reputation. While fixed unobserved heterogeneity across researchers in aspects such as personal characteristics, PhD training or ability is absorbed by the individual dummies includes in the panel structure, it is possible that changes in reputation influence future output. This is relevant because there is a notion that the distribution of recognition in science is influenced by a "class structure" (Merton, 1968) that is skewed in a way that favor of those researchers who already have reputation. In fact, there is some evidence suggesting that scientific reputation has an effect on the expected ability to secure research grant funding (Arora et al 1998 and Arora and Gambardella 1998). Moreover, it is reasonable to expect that one or several of the measures of network structure or relations correlate with reputation. Thus, in studying scientific productivity, a control for reputation helps to make sure that greater levels of output would are indeed resulting from the social capital the researcher is building and are not confounded effects with the reputation developed by the researchers.

Different measures were used to control for reputation: 1) the number of past publications in a four-year window (t-1)[9], 2) the number of single-authored papers that the researcher has published in the past three years, and 3) the international visibility by counting the number of articles with a foreign address in the same three-year window. We believe that these 3 indicators are a reasonable measures of reputation since most universities used at least one of them to give raises and promotions.

Finally, we also considered the variables past publications or past cites because including a lagged dependent variable helps to control for unobserved individual time variant variables and for other potentially important, but potentially omitted predictors (Greene 2000).

4. Results

Table 1 shows the descriptive statistics for the variables. Our sample of researchers has a mean of 1.8 publications per two years with a standard deviation of 3 publications. Each of these publications received on average 6.4 citations in the next four years with standard deviations of 18.4. The mean number of direct ties is 6, and the strength of those ties is 1.5. In McFadyen and Canella (2004) study, the number of relations is substantially larger (46.5), while the strength of those relations is lower (1.33). The mean of the structural holes variable is 0.41, and the mean for density is 82.8,.. The mean of centrality is 0.34 and since the mean of the external-internal index is negative, it is observed that scientists tend to collaborate more with peers within the same area of knowledge.

[9] This measure was done in a 4-year window, instead of 3 to reduce the correlation between direct ties and past publications.

Table 1. Descriptive Statistics

Variable	Mean	Std. Dev.
Pubs	1.7642	3.0050
Cites	6.3630	18.3926
Directties	6.2745	6.8184
Strengthdt	1.4534	0.6599
Structholes	0.4120	0.2952
Density	82.7719	30.1232
Eigenvec	0.3438	4.0750
E-I Index	-0.3538	0.7259
Past pubs	2.5619	4.6152
Past cites	6.9725	22.1873
Sing auth	0.1694	0.6775
Int art	0.7580	2.0038
Observations		
N=15336, n=1704, T=9		

Table 2 shows the correlation of our variables. As can be seen there are high correlations among some variables. In particular, there is a high correlation between structural holes and density, and the number of direct ties and past publications. Although this high correlation among some of the variables used could cause a multicollinearity problem, unbiased estimators are still produced, and this problem only would increase the variances for the collinear variables (Kennedy 1998). Therefore, even if collinearity problems exist, the results are still unbiased and finding significance produces valid results.

Table 2. Correlations

	Pubs	Cites	Directties	Strengthdt	Structholes	Density	Eigenvec	E-I Index	Past pubs	Past cites	Sing auth	Int art
Pubs	1											
Cites	0.6952	1										
Directties	0.5469	0.4247	1									
Strengthdt	0.2619	0.1824	0.1831	1								
Structholes	0.3633	0.2571	0.6428	0.0825	1							
Density	-0.401	-0.2675	-0.5448	-0.0978	-0.7266	1						
Eigenvec	0.2647	0.3519	0.326	0.1385	0.0957	-0.0946	1					
E-I Index	0.0051	-0.0697	0.0681	-0.0226	0.0598	-0.0471	-0.0539	1				
Past pubs	0.6547	0.5229	0.7656	0.4481	0.5007	-0.5735	0.3661	0.0198	1			
Past cites	0.5458	0.652	0.5347	0.3147	0.3203	-0.3507	0.4617	-0.0759	0.6946	1		
Sing auth	0.2017	0.1495	0.0119	0.0136	0.0366	-0.1433	0.0255	-0.0063	0.2924	0.2	1	
Int art	0.4779	0.5592	0.5659	0.2891	0.3405	-0.3732	0.5036	-0.0664	0.674	0.746	0.1259	1

The results of the regressions are shown in Table 3. In models 1 to 7 the dependent variable is the straight count of publications. Models 8 to 14 show the results of the regressions where the dependent variable is the number of cites in the next four years. Given that the last year of information is 2002; only periods 1 to 7 were included. Models 2 to 7 for publications and 9 to 14 for cites show the isolated effect of each network variable to analyze which of those are unstable when other variables are included.

Table 3. Models

	Model 1 Neg Bin Fixed Eff			Model 2 Neg Bin Fixed Eff			Model 3 Neg Bin Fixed Eff			Model 4 Neg Bin Fixed Eff		
Pubs	Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.	
Directties	0.0052	0.0025	*	0.0068	0.0022	***						
Strengthdt	-0.0182	0.0185					-0.0358	0.0175	*			
Structholes	-0.1670	0.0694	*							0.1202	0.0489	**
Density	-0.0028	0.0005	***									
Eigen	0.0061	0.0018	***									
E-I Index	0.0695	0.0251	**									
Past pubs	-0.0061	0.0038		-0.0038	0.0035		0.0037	0.0029		0.0004	0.0029	
Sing auth	0.0355	0.0129	**	0.0353	0.0128	**	0.0270	0.0126	*	0.0298	0.0126	*
Int art	0.0004	0.0052		0.0000	0.0052		0.0021	0.0051		0.0018	0.0052	
Period 8	0.0095	0.0262		0.0074	0.0262		0.0058	0.0262		0.0108	0.0263	
Period 7	-0.1278	0.0299	***	-0.1349	0.0297	***	-0.1392	0.0297	***	-0.1313	0.0299	***
Period 6	-0.1387	0.0336	***	-0.1361	0.0333	***	-0.1426	0.0332	***	-0.1315	0.0336	***
Period 5	-0.3570	0.0400	***	-0.3636	0.0393	***	-0.3708	0.0392	***	-0.3551	0.0400	***
Period 4	-0.5036	0.0450	***	-0.5155	0.0441	***	-0.5264	0.0440	***	-0.5053	0.0449	***
Period 3	-0.7505	0.0530	***	-0.7480	0.0508	***	-0.7583	0.0507	***	-0.7288	0.0524	***
Period 2	-0.6736	0.0554	***	-0.6868	0.0543	***	-0.6989	0.0543	***	-0.6735	0.0553	***
Period 1	-0.6940	0.0598	***	-0.7189	0.0587	***	-0.7334	0.0586	***	-0.7034	0.0599	***
Cons	2.3366	0.0968	***	2.0017	0.0700	***	2.0523	0.0726	***	1.9509	0.0736	***

	Model 5 Neg Bin Fixed Eff			Model 6 Neg Bin Fixed Eff			Model 7 Neg Bin Fixed Eff		
Pubs	Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.	
Directties									
Strengthdt									
Structholes									
Density	-0.0050	0.0004	***						
Eigen				0.0058	0.0018	***			
E-I Index							0.0776	0.0249	***
Past pubs	-0.0065	0.0033	*	0.0013	0.0028		0.0012	0.0028	
Sing auth	0.0897	0.0119	***	0.0323	0.0126	**	0.0284	0.0126	*
Int art	0.0118	0.0056	*	0.0004	0.0051		0.0030	0.0051	
Period 8	0.0121	0.0270		0.0045	0.0263		0.0104	0.0262	
Period 7	-0.1993	0.0297	***	-0.1450	0.0298	***	-0.1349	0.0297	***
Period 6	-0.3662	0.0326	***	-0.1495	0.0333	***	-0.1359	0.0333	***
Period 5	-0.7091	0.0379	***	-0.3843	0.0393	***	-0.3639	0.0393	***
Period 4	-1.0394	0.0429	***	-0.5396	0.0441	***	-0.5146	0.0441	***
Period 3	-1.3844	0.0483	***	-0.7715	0.0507	***	-0.7463	0.0508	***
Period 2	-1.5100	0.0518	***	-0.7127	0.0544	***	-0.6860	0.0545	***
Period 1	-1.7995	0.0581	***	-0.7427	0.0585	***	-0.7137	0.0589	***
Cons	1.5769	0.0539	***	2.0209	0.0696	***	2.0452	0.0707	***

*** significant at 0.1%, ** significant at 1%, * significant at 5%

Considering that the use of panel data could entail autocorrelation problems among variables, Wooldridge (2002 pp 282-283) tests were performed to be certain that these analyses do not have this potential bias. The results reject the possibility of autocorrelation.

Table 3. Models (continuation)

	Model 8 Neg Bin Fixed Eff			Model 9 Neg Bin Fixed Eff			Model 10 Neg Bin Fixed Eff			Model 11 Neg Bin Fixed Eff		
Cites	Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.	
Directties	0.0180	0.0045	***	0.0250	0.0037	***						
Strengthdt	0.1035	0.0274	***				0.0800	0.0271	***			
Structholes	-0.0379	0.1121								0.4100	0.0805	***
Density	-0.0041	0.0009	***									
Eigen	0.0036	0.0034										
E-I Index	-0.0253	0.0373										
Past cites	0.0013	0.0008		0.0018	0.0008	*	0.0021	0.0008	**	0.0022	0.0008	**
Sing auth	0.0859	0.0210	***	0.0875	0.0209	***	0.0805	0.0215	***	0.0809	0.0212	***
Int art	-0.0088	0.0116		0.0007	0.0113		0.0164	0.0109		0.0141	0.0110	
Period 6	-0.0093	0.0481		-0.0238	0.0478		-0.0515	0.0476		-0.0332	0.0477	
Period 5	-0.1276	0.0570	*	-0.1533	0.0564	**	-0.2128	0.0556	***	-0.1604	0.0566	**
Period 4	-0.2837	0.0636	***	-0.3198	0.0625	***	-0.3870	0.0615	***	-0.3246	0.0628	***
Period 3	-0.3916	0.0710	***	-0.4136	0.0680	***	-0.4868	0.0669	***	-0.3946	0.0694	***
Period 2	-0.5469	0.0778	***	-0.5938	0.0764	***	-0.6702	0.0752	***	-0.6041	0.0764	***
Period 1	-0.3580	0.0815	***	-0.4106	0.0798	***	-0.5033	0.0782	***	-0.4227	0.0800	***
Cons	-0.2657	0.1174	*	-0.3975	0.0588	***	-0.3460	0.0661	***	-0.4435	0.0676	***

	Model 12 Neg Bin Fixed Eff			Model 13 Neg Bin Fixed Eff			Model 14 Neg Bin Fixed Eff		
Cites	Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.	
Directties									
Strengthdt									
Structholes									
Density	-0.0081	0.0007	***						
Eigen				0.0023	0.0035				
E-I Index							0.0029	0.0365	
Past cites	0.0399	0.0048	***	0.0022	0.0008	**	0.0023	0.0008	***
Sing auth	0.0922	0.0171	***	0.0816	0.0216	***	0.0798	0.0215	***
Int art	-0.0128	0.0092		0.0185	0.0112		0.0202	0.0109	
Period 6	-0.1765	0.0420	***	-0.0650	0.0477		-0.0616	0.0475	
Period 5	-0.4167	0.0475	***	-0.2212	0.0559	***	-0.2166	0.0557	***
Period 4	-0.7755	0.0533	***	-0.4000	0.0619	***	-0.3941	0.0617	***
Period 3	-1.0192	0.0573	***	-0.4956	0.0673	***	-0.4896	0.0673	***
Period 2	-1.3001	0.0640	***	-0.6903	0.0755	***	-0.6837	0.0756	***
Period 1	-1.4239	0.0683	***	-0.5250	0.0782	***	-0.5196	0.0786	***
Cons	-0.2111	0.0658	***	-0.2191	0.0526	***	-0.2254	0.0526	***

*** significant at 0.1%, ** significant at 1%, * significant at 5%

The results of Model 1 show that direct ties, centrality and external-index enhance productivity, confirming hypothesis 1, 5 and 6. Also, hypothesis 3 is rejected since dense networks are found to negatively affect future productivity. Similarly, structural holes are also found to affect negatively the outcomes of researchers, contradicting hypothesis 4. The result for strength of direct ties is not significant. When looking at control variables, only the single-authored papers variable is found to be significant, affecting positively the productivity of researchers[10].

Models 2 to 7 show the results of the regressions when each of the individual network variable is considered. As can be seen the sign and the significance of direct ties, density, centrality and external-internal index remain equal when the complete set of variables are included in the regression as well as when each of the individual variables are considered. In the case of the strength of ties, the sign remains but it lost significance when the other variables are included. The only variable that changes sign and significance is structural holes. An interpretation of these results is discussed in the next section.

The results of Model 8 show that, when adjusting for quality, direct ties and density produce the same effect that when looking at publications; direct ties affects positively the number of citations, and researchers embedded in dense networks tend to receive less citations. For this case, the strength of direct ties, which was not significant for publications, is positive, contradicting hypothesis 2 and suggesting that frequent interactions could produce higher quality papers. The rest of the network variables, structural holes, centrality and external-internal index are not significant. As in the case of publications, the only control variable that is significant and affects positively the number of citations is single author papers.

Models 9 to 14 show the results of the regressions considering each of the individual network variables. The variables that are stable, this is that do not change sign and significance when all the set of variables are considered and when they are included by itself, are density, strength of direct ties and density. Centrality is also stable although it is not significant in either of the regressions. The external-internal index is also not significant in either of the cases but it changes sign. As in the case of publications, the structural holes variable changes sign and significance. In this case, it is positive and

[10] Similar regressions were run considering only each of the individual control variables. The sign and and significance of the network variables are the same, thus we decided to report the result considering all control variables.

significant when it is considered in isolation, and it is negative and not significant when all the set of network variables are incorporated in the regression. A discussion is presented in the next section.

Finally, we test for potential diminishing returns in the number of direct ties and strength of direct ties to compare the results with those of McFadyen and Cannella (2004). However, in both of our measures of productivity, straight count of publications and number of cites in the next four years, no evidence of diminishing returns in the number of direct ties or the strength of those ties is found. The results of these regressions are shown in an appendix.

In conclusion, strong evidence is found to accept hypothesis 1 and 3, the higher the number of direct ties, the higher the productivity and the quality of that productivity; and researchers embedded in dense networks tend to publish less papers and receive less citations.

5. Discussion

Evidence is found that controls for different aspects of social capital affects the result of the impact of several established network variables. In particular, the analysis suggests that the structural holes variable changes sign and significance when other variables are included. These results suggest that to have a clear understanding of social capital measures and their impact on productivity, it is critical to control for the various dimensions of social capital, as well as for unobserved individual heterogeneity. Most existing studies have found structural holes to be positively related to performance (Burt 1992; Burt 2001) and, in fact, this argument has echoes from the past (Krackhardt 1999), going back to Lin Freeman’s work on betweenness centrality (Freeman 1979) and Granovetter’s work on the strength of weak ties (Granovetter 1973). However, most studies related to social capital do not control for unobserved heterogeneity, since they relied mainly on cross section data (Burt 2000; Soda and Zaheer 2004). It is likely that both the differences in the number of publications among researchers and the differences in their network variables result from unobserved stable characteristics of the researchers, such as overall intellectual ability or how sociable they are. Since in a cross section it is not possible to control for these unobserved characteristics, results may be biased. Another issue relates to the need to control for complementary and competing aspects of a network when trying to identify their individual impact on productivity. [As was seen, the structural holes variable changes sign and significance when other network variables are taken into account. Other important aspects are that we undertake two substantial shortcomings of most of the research done in social network analysis (Gulati 1995; Newman 2001). We are considering a complete network, and our measures of social capital are objective, since they are not sensitive to the subjective bias that comes with interviews.

Another important aspect that could be confounded with the effects of changes in networks variables on the productivity of individuals is location. In countries with researchers' mobility, changes of institution would be driven by past success, as well the expectation of increasing productivity. But when they move, the new environment allows dramatic changes in the environment to which they are exposed, affecting their collaboration networks. This could lead to reverse causality of the impact of particular network variables on researcher productivity. But in Mexico there is almost no mobility and there are no formal rules of collaboration. As a result, one could argue that this study enables better control for these issues than an alternative one using comparative data from the U.S.

Thus, evidence is found that structural holes variable is not necessarily the best measure of social capital in academia. On the contrary, other measures appear to have greater impact on the production of new knowledge. In particular, researchers with high number of direct ties, who are part of sparse network, who are central (closer to other central colleagues) and who collaborate with researchers in other disciplines reach higher productivity.

The fact that density appears to affect future productivity negatively could be an important issue for university administrators and science policy makers. They can design policies to dissuade dense networks. For example, inbreeding and the lack of mobility in general could boost dense network since people could tend to cluster with the same people, and this trend may affect negatively the creation of new knowledge Horta et al (2007) find that inbred faculty are less productive, more centered in their in their own institutions and less open to the rest of the scientific community. [].

Another important result is that the number of direct ties is positively related to future productivity, thus researchers could try to increase their academic relationships, and be more open to the scientific world. Since there is no evidence of diminishing returns, the resources needed to create and maintain those relations seem not to be a concern (Granovetter, 1973). Additionally, since centrality also play a key role in productivity, researchers could also procure to be team leaders. One common way is by being active in teaching activities, specifically by having PhD students.

This study also shows that having researchers collaborating across discipline boundaries is associated with better productivity. This could be associated with the idea that nowadays, researchers tend to investigate more complex problems (Adams et al 2005),and interdisciplinary and multidisciplinary collaborations are needed. Thus, researchers may try to increase their academic relationships with people from other disciplines not only to study multifaceted problems but also to have diverse views of tackle academic topics.

This study has also some limitations. First, co-publication is not an exhaustive measure of collaboration; there are more products of collaboration than joint publications in ISI. In addition, it is possible that researchers give co-authorship to some scientists just because they work at the same laboratory or they share some equipment (even if they do not contribute much to the work), so that the analysis of co-authorship might not reflect the actual relationships in academic networks (Stephan 1996).

No less important is the fact that, while social capital has a “powerful and intuitive appeal”, (Dasgupta 2005), it is very difficult to measure because it has many and varied components, and some of them are intangible. This study has not considered all the relationships that someone could have and that could affect, to some degree, the creation of new knowledge. For example, interaction with students could be a valuable source of novel information, and that interaction might not always end in a joint publication. Moreover, only quantitative variables to proxy for social capital are used, without including the wide range of social phenomena than involves human relationships. Also important is the fact that we are not taking into account the pathways by which networks are formed and the motivations behind their formation. These are important factors that could lead to a better design of networks that promote the creation of new knowledge.

Finally, it is acknowledged that social capital affects in a different way the creation of knowledge depending on the specific area of knowledge. Although we tried to minimize this problem by considering one broad area of knowledge, the main conclusions of this study may not equally apply to all fields of knowledge

6. Conclusions

This study looks at how networks condition the creation of new knowledge. One of the most important results is that controlling for reputation and unobserved heterogeneity has a major impact on the existing notions regarding the effects of network variables on performance. The results show that dense networks affect negatively the creation of new knowledge. In addition, they show that interdisciplinary work increases research output. It is also found that the position in a network is a critical aspect, such that researchers who are central tend to create more knowledge, even when controlling for past performance and reputation. Finally, findings suggest that the number of direct ties impacts positively future productivity.

Another critical conclusion of this study is the finding that structural holes appear not to be the best measure of social capital. On the contrary, this variable changes sign and significance when other variables and controls for unobserved heterogeneity are included. This result contradicts what has been considered one of the most important advances in the social network literature (Krackhardt 1999).

Although the study has limitations,, we think that the contributions of this paper are of significant importance, not only to the literature of social network analysis, but also for a better design of science policy.

Acknowledgments

The authors are grateful to Ashish Arora, Robert Lowe and Granger Morgan for their insightful comments and suggestions. An earlier version of this paper was presented at the Academy of Management 2006. Support from Asociación Mexicana de Cultura and the Mexican Council for Science and Technology (Conacyt) is gratefully acknowledged.

Appendix

Table 4. Diminishing Returns in the Number and Strength of Direct Ties

	Neg Bin Fixed Eff				Neg Bin Fixed Eff		
Pubs	Coef.	Std. Err.		Cites	Coef.	Std. Err.	
Directties	0.0032	0.0035477		Directties	0.0276	0.0089	***
Directtiessq	0.0000	0.0000265		Directtiessq	-0.0002	0.0002	
Strengthdt	-0.0231	0.0370987		Strengthdt	0.1741	0.0599	***
Strengthdtsq	0.0008	0.0054809		Strengthdtsq	-0.0113	0.0087	
Structholes	-0.1533	0.0715933	*	Structholes	-0.0866	0.1180	
Density	-0.0028	0.0005462	***	Density	-0.0039	0.0009	***
Eigen	0.0061	0.001826	***	Eigen	0.0042	0.0035	
E-I Index	0.0700	0.0251183	**	E-I Index	-0.0286	0.0374	
Past pubs	-0.0059	0.0038434		Past Cites	0.0012	0.0008	
Sing auth	0.0352	0.0129623	**	Sing auth	0.0865	0.0209	***
Int art	0.0009	0.005212		Int art	-0.0110	0.0117	
Period 8	0.0096	0.0262168					
Period 7	-0.1285	0.0299381	***				
Period 6	-0.1404	0.0336793	***	Period 6	-0.0065	0.0481	
Period 5	-0.3596	0.0401988	***	Period 5	-0.1173	0.0572	*
Period 4	-0.5059	0.0451251	***	Period 4	-0.2795	0.0636	***
Period 3	-0.7513	0.0529841	***	Period 3	-0.3941	0.0711	***
Period 2	-0.6764	0.0554874	***	Period 2	-0.5405	0.0778	***
Period 1	-0.6968	0.0599205	***	Period 1	-0.3493	0.0815	***
Const	2.3472	0.1019951	***	Const	-0.3737	0.1321	**

*** significant at 0.01%, ** significant at 1%, * significant at 5%

References

Adams, J.D., Black, G.C., Clemmons, J.R. and Stephan, P.E. 2005."Scientific Teams and Institutional Collaborations: Evidence from U.S. Universities, 1981-1999", *Research Policy*. **34** 259-285.

Ahuja, G. 2000. "Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study", *Administrative Science Quarterly*. **45** 425-455.

Arora, A. and Gambardella, A. 1998. "The Impact of NSF Support for Basic Research in Economics". *Working paper*. 1-24.

Arora, A., David, P, & Gambardella, A. 1998. "Reputation and Competence in Publicly Funded Science", *Annales d'Economie et de Statistiques*. **49/50** 163-196.

Arora, A., Gambardella, A. 1990. "Complementarity and External Linkages: the Strategies of the Large Firms in Biotechnology," *Journal of Industrial Economics*. **38** 361-379.

Barnett, A.H., Ault, R.W., Kaserman, D.L. 1988. "The Rising Evidence of Coauthorship in Economics: Further Evidence", *The Review of Economics and Statistics*. **70**, **3** 539-543.

Bonacich, P.B. 1972. "Technique for Analyzing Overlapping Memberships in: H. Costner, Ed., *Sociological Methodology*. Jossey-Bass, San Francisco, 176-185.

Bogartti, S.P. 1995. "Centrality and AIDS", *Connections*. **18, 1** 112-115.

Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet 6 for Windows. Harvard: Analytic Technologies.

Burt, R.S. 1992. *Structural Holes: The Social Structure of Competition*. Cambridge: Harvard University Press.

Burt, R.S. 2001. "Structural Holes versus Network Closure as Social Capital" *in Social Capital: Theory and Research*, edited by Nan Lin, Karen S. Cook and R.S. Burt. Aldine de Gruyter.

Coleman, J.S. 1998. "Social Capital in the Creation of Human Capital", *American Journal of*

Sociology. **94** 95-120.

CONACYT. 2003. *Indicadores de Actividades Científicas y Tecnológicas 2002*, Mexico.

Cross R. and Cumming, J.N. 2004. "Tie and Network Correlates of Individual Performance in Knowledge-Intensive Work", *Academy of Management Journal*. **7, 6** 928-937.

Dasgupta, P. 2005. "Economics of Social Capital", *The Economic Record*. **81** s1-s21.

Green, W.H. 2000. *Econometric Analysis* (4th ed.), Upper Saddle River, NJ, Prentice Hall.

Gil Antón, M. (coordinador). 1994. *Los Rasgos de la Diversidad: Un Estudio Sobre los Académicos Mexicanos*. Universidad Autónoma Metropolitana.

Gonzalez-Brambila, C. and Veloso, F. 2004. "The Determinants of Research Productivity: A Study of Mexican Researchers. *Working paper*.

Hausman, J., Bronwyn, H. and Griliches, Z. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship", *Econometrica*. **52** 909-938.

Inkpen, A.C. and Tsang, E.W.K. 2005. "Social Capital, Networks, and Knowledge Transfer", *Academy of Management Review*. **30**, **1** 146-165.

Institute of Scientific Information, ISI (2003, April 18). *National Citation Report for Mexico*. Articles from ISI database years 1981-2002 cited in the period 1981-2002. Citing and cited papers are included. Project 2487.

Jaffe, A., Trajtenberg, M., and Henderson, R. 1993. "Geographic Localization of Knowledge Spillovers as Evidenced from Patent Citations", *Quarterly Journal of Economics*. **108** 557-598.

Katz, J.S. 1994. "Geographical Proximity of Scientific Collaboration", *Scientometrics*. **31, 1** 31-43.

Krackhardt, D. 1999. "The Ties that Torture: Simmelian Tie Analysis of Organizations", *Research in the Sociology of Organizations*. **16** 183-210.

Levin, S. and Stephan, P. 1991. Research Productivity Over the Life Cycle: Evidence for

Academic Scientists, *The American Economic Review*. **81** 114-132.

Mairesse, J. and Turner, L. 2002. “Measurement and Explanation of the Intensity of Co-publication in Scientific Research: An Analysis at the Laboratory Level”. *Working paper*.

McFadyen, M.A. and Cannella, A. 2004. “Social Capital and Knowledge Creation: Diminishing Returns of the Number and Strength of Exchange Relationships”, *Academy of Management Journal*. **47, 5** 735-746.

Melin, G. 1999. “Impact of National Size on Research Collaboration: A Comparison Between Northern European and American Universities”, *Scientometrics*. **46** 161-170.

Melin, G., and Persson, O. 1996. “Studying Research Collaborations using Co-authorship”, *Scientometrics*. **36** 363-377.

Merton, R. 1968. “The Matthew Effect in Science. The Reward and Communication Systems of Science are Considered”, *Science*. **159** (3810) 56-63.

Nahapiet, J. and Ghoshal, S. 1998. “Social Capital, Intellectual Capital, and the Organizational Advantage”, *The Academy of Management Review*. **23, 2** 242-266.

Nonaka, I. 1994. “A Dynamic Theory of Organizational Knowledge Creation”, *Organization Science*. **5, 1** 14-37.

Stephan, P. 1996. “The Economics of Science”, *Journal of Economic Literature*, **XXXIV,** 1199-1235.

Stephan, P. and Levin, S. 1997. “The Critical Importance of Careers in Collaborative Scientific Research”, *Revue d’Economie Industrielle*. **79** 45-61.

Turner, L. & Mairesse, J. 2003. “Explaining Individual Productivity Differences in Scientific Research Productivity: How important are Institutional and Individual Determinants? An Econometric Analysis of the Publications of French CNRS Physicists in Condensed Matter (1980-1997)”, *Submitted to Annales d’Economie et de Statistiques* for the special issue in honor of Zvi Griliches.

Wang, Y., Wu, Y., Pan, Y., Ma, Z., Rousseau, R. 2005. "Scientific Collaboration in China as Reflected in Co-authorship", *Scientometrics*. **62, 2** 183-198.

Wasserman, S., and Faust, K. 1994. *Social Network Analysis: Methods and Applications*. Cambridge, England. Cambridge University Press.

Wooldridge, J.M. 2002. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.