Research Policy 42 (2013) 1555–1567

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

The impact of network embeddedness on research output

Claudia N. Gonzalez-Brambila[a,*], Francisco M. Veloso[b,c], David Krackhardt[d]

[a] Business School, Instituto Tecnologico Autonomo de Mexico, Mexico, D.F. 01080, Mexico
[b] Católica Lisbon School of Business and Economics, 1649-023 Lisbon, Portugal
[c] Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213, United States
[d] Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213, United States

ARTICLE INFO

Article history:
Received 5 May 2010
Received in revised form 24 June 2013
Accepted 23 July 2013
Available online 24 August 2013

Keywords:
Network embeddedness
Research output and impact
Mexico

ABSTRACT

This paper aims to further our understanding of how embeddedness affects the research output and impact of scientists. The analysis uses an extensive panel data that allows an analysis of within person variation over time. It explores the simultaneous effects of different dimensions of network embeddedness over time at individual level. These include the establishment of direct ties, the strengths of these ties, as well as the density, structural holes, centrality, and cross-disciplinary links. Results suggest that the network dynamics behind the generation of quality output contrasts dramatically with that of quantity. We find that the relational dimension of scientists matters for quality, but not for output, while cognitive dimensions have the opposite effect, helping output, while being indifferent toward impact. The structural dimension of the network is the only area where there is some degree of convergence between output quantity and quality; here, we find a prevalence for the role of brokerage over cohesion. It concludes by discussing implications for both network research and science policy.

1. Introduction

It is widely assumed among scientists that research collaboration is desirable. Beaver and Rosen (1978) identified many motives for collaboration: to increase productivity, access special equipment and facilities, access special skills, access unique materials, visibility, efficiency in use of time, to gain experience, to avoid competition, and spatial propinquity, among many others. This empirical study provides a good idea of the range of motives leading researchers to prefer collaboration over being sole authors.

There is a variety of empirical evidence showing that collaboration is beneficial. pioneer work on research productivity found a strong relationship between collaboration and scientific productivity. Price and Beaver (1966) also found that the most productive researchers were also the most collaborative, a result confirmed by Zuckerman (1967) a year later. Furthermore, it has been shown that there is a positive relation between co-authorship and article's number of cites (Katz and Hicks, 1997; Glanzel and Schubert, 2001; Wuchty et al., 2007). 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 and international collaboration and geographic dispersion of team members (Adams et al., 2005; Stephan and Levin, 1997; Wang et al., 2005; Melin, 1999; Melin and Persson, 1996; Barnett et al., 1988; Katz, 1994; He et al., 2009; Singh, 2007; Lee and Bozeman, 2005). In addition, a complementary line of research has looked at the relationship between network structure and innovation (Ahuja, 2000; Gilsing et al., 2008).

While existing studies looking at the impact of collaboration on scientific productivity have provided a variety of important insights, we still have a limited understanding of how network embeddedness established through collaborations conditions scientific output and impact. Network embeddedness dimensions include relational embeddedness, structural embeddedness and cognitive embeddedness (Nahapiet and Ghoshal, 1998). The relational dimension breadth and depth of personal relationships scientists develop with each other through the publishing effort (Granovetter, 1992). The structural dimension refers to the pattern of connections between researchers, who do they reach and how they reach them (Burt, 1992). Finally, the cognitive dimension refers to those resources that provide shared interpretations, languages and codes. In the context of the scientific endeavor, this will mean the fields of knowledge in which researchers develop their work.

Several issues contribute to our limited understanding of how network embeddedness conditions output and impact. The first and most significant issue is the controversy between brokerage and closure as drivers of performance. Existing theories behind the

* Corresponding author.
E-mail addresses: cgonzalez@itam.mx (C.N. Gonzalez-Brambila), fveloso@ucp.pt (F.M. Veloso), krack@andrew.cmu.edu (D. Krackhardt).

0048-7333/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.respol.2013.07.008

importance of network embeddedness for productivity have been sometimes contradictory. On one camp, authors have made a case for brokerage. The idea is that open social structures with many structural holes can connect new members in different clusters and get access to new information, which will benefit performance. For example, Burt (2004) finds that higher compensations, positive performance evaluations or promotions are disproportionately in the hands of managers whose networks are rich in structural holes. A contrasting perspective is one of closure: actors in tighter embedded networks obtain more coordination, trust each other, and develop better communication skills, which lead to superior results (Coleman, 1988). Despite a growing number of empirical studies (e.g. Burt, 1992, 2001; Ahuja, 2000; Fleming et al., 2007; Reagans and McEvily, 2008; Obstfeld, 2005), the relative importance of each dimension is far from being settled.

A second issue is the interrelation between the various dimensions of network embeddedness. Existing research suggests that the effect of one dimension of network embeddedness can temperate the effect of another. For example Fleming et al. (2007) suggest that greater network density, combined with a larger set of ties, can increase one's ability to generate novel knowledge because density facilitates trust among network participants, increasing their willingness to share valuable knowledge. In the context of science, it is well known that high quality scientists, not only publish more, but also form more collaborative relationships (Zuckerman, 1967; Zucker and Darby, 2006). Similarly, it is straightforward to conceive that superior past performance can lead to advantages in forming network ties, helping network brokers to establish new, non-redundant ties (Burt, 1992). As a result, to properly understand the role of the various network dimensions and conclude on the relative contribution of each of them for performance, the influence of all variables should be taken into consideration. The need for such integrated perspective is recognized in a recent review (Phelps et al., 2012), which notes the lack of consistency in results, and the opportunity for careful work looking at "[. . .] mechanisms linking observed knowledge network elements and knowledge outcomes and moderators of these mechanisms" [p. 1136].

The third dimension that we will consider, which we believe can influence the role that network embeddedness play in performance, is the actual measurement of performance, especially when contrasting output and impact. Most prior studies use one measure of productivity in the analysis (Burt, 2001; Gilsing et al., 2008; Ahuja, 2000) and thus have not quite explored this aspect. This issue has certainly not been widely explored in an environment like academia (McFadyen and Cannella, 2004; He et al., 2009; Singh, 2007; McFadyen et al., 2009). We expect that a more nuanced perspective on how one measures knowledge output can help contextualize and better explain the influence of the various dimensions of embeddedness.

This paper explores the simultaneous effects of different dimensions of network embeddedness on research output and impact by isolating the marginal effects of the characteristics of the ego-network from the individual effects of the agents within the collaboration network, both in their observed (education, gender, etc.) and unobserved characteristics. A crucial characteristic of this research is the use of an extensive panel data, which allows us to explore the effect of network embeddedness on different outcomes variables and to do the analysis considering within person variation over time. This is important for several reasons. First, existing studies in the network literature often look at differences between people to identify how changes in network characteristics may relate to some desired outcome variable (e.g. Burt, 2001; Reagans and McEvily, 2008). While these results are certainly interesting, they offer a limited insight into the questions of how changes in a particular dimension of social network embeddedness around an individual conditions relevant outcomes. The critical issue is that it is reasonable to consider that individual characteristics unobserved by the econometrician, for example the intrinsic quality of an individual as measured by intellectual ability, might be correlated with particular network dimensions as well as with the outcome variable, say output or productivity (Singh, 2007; He et al., 2009). Thus, controlling for individual specific heterogeneity may yield different results when compared to an assessment based on differences across individuals or organizations (Fleming et al., 2007). Second, time varying individual characteristics could also be confounded with critical network variables, again leading to divergent results. Third, one may actually have different theoretical predictions for a setting that looks at differences between individuals versus one that considers within person variation. For example, one could consider that the strength of the ties of an individual in a network relates only to intrinsic individual ability, such that a cross sectional analysis would find a correlation between this variable and performance, but a time varying analysis would not support such a relation. We address all these issues by employing fixed effects models at the level of the individual scientists. Finally, as noted above, it is possible that the nature of the networks produces different effects depending on the performance variable that is taken into consideration. Our dataset allows us to consider different outcomes and thus address this issue.

Overall, we aim to contribute to the literature on research productivity by advancing our knowledge of the specific mechanisms that make research collaboration work, and how they impact individual output and impact. We also wish to contribute to the network literature by improving our understanding of the interactions of the different dimensions of network embeddedness, and how they jointly condition performance.

The structure of the paper is as follows. The next section provides some theoretical background and the hypotheses. 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 section has the discussion and conclusions.

2. Theoretical background and hypothesis

Since Lotka's (1926) pioneer analysis on scientific productivity, several studies have found a positive relationship between collaboration and scientific productivity. De Solla Price and Beaver (1966) conclude that the most prolific researchers are also those that collaborate the most. Zuckerman (1967) showed that Nobel Laureates published more and were more apt to collaborate than a matched sample of scientists. Other research also found a positive association between collaboration and the impact of scientific publications (Narin et al., 1991; Katz and Hicks, 1997; Glanzel and Schubert, 2001). These studies have related co-authorship, particularly international co-authorship, and the number of citations as a measure of quality and impact. He et al. (2009) used a longitudinal study of 65 biomedical scientists to show that within university collaboration and international collaboration are positively related to an article's quality, and that international collaboration is positively related to a scientist's future productivity. In another study, Wuchty et al. (2007) find that coauthored papers tend to receive more citations than sole-authored papers.

Considering that the production of scientific knowledge is deeply embedded in social structures and practices among scientists (Katz and Martin, 1997), there have been few studies which have used social network analysis to explore the characteristics of social embeddedness of scientific collaboration and their impact on performance. McFadyen and Cannella (2004) use a sample of publications of 173 biomedical scientists from 2 universities to test the relationship between network ties and knowledge creation. They find that superior knowledge creation is associated with 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 this study provides important insights, it also has some limitations. As acknowledged by the authors, the sample is small and only considers biomedical scientists, so it may not be representative. The study is also limited because it focuses only on direct ties and their strength, ignoring other important aspects of social embeddedness. Singh (2007) uses a longitudinal study of worldwide scientists in Biotechnology and Applied Microbiology to explore the impact of external collaboration across national, organizational or institutional boundaries. In his study, he explores not only the impact of the collaborative article itself, but also the effect in the future productivity of collaborating scientists. Moreover, he analyses the influence of the number and intensity of interpersonal ties, as well as the benefits from structural holes in the creation of new scientific knowledge. His findings confirm that external collaboration improves future productivity. He also shows that the three characteristics of a scientist's ego-centric network have a positive effect on productivity. Despite the important findings of this study, it does not take into consideration other aspects of social embeddedness that could be relevant, such as density and position in the ego-network, and the similarities among the authors' field of knowledge, which are important dimensions of network embeddedness.

Previous work raises the question of how does embeddedness in academia influence the productivity and impact of a scientist. Up to now, there is no compelling evidence of what kind of network embeddedness enhances the creation of new knowledge. Some 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.

In the network structure literature, two opposing versions of the theory 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 relationships.[1] The idea is that these actors have advantages because they can get access and transmit new information sooner than actors on the periphery.

Most 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 Cumming, 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 Cumming, 2004). We argue that to better understand the social relationship in academia, it is necessary to analyze the different dimensions of social capital: relational, structural and cognitive (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 the most critical variables that can potentially influence the access to new information.

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

The first relational dimension we consider is the number of direct relationships that a given actor has at a given time. These are expected to stimulate combination and exchange of resources that exist within the relationships (Nahapiet and Ghoshal, 1998) 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, and resources people have access to, and therefore enhance their ability to address the complex problems in the form of advanced research (McFadyen and Cannella, 2004; Reagans and McEvily, 2003; Ahuja, 2000). Hence, we expect,

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

A second relational dimension is the strength of the relationships. As stated before, research collaboration could be beneficial, but this cooperation also entails various costs. For example, costs of finding the right partner, costs of organizing and distributing teamwork, costs of developing a shared understanding, trust, reciprocity and the transfer of high quality information and tacit knowledge, among other costs (Reagans and McEvily, 2003; Rowley et al., 2000; Gulati, 1995; Larson, 1992; Inkpen and Tsang, 2005). To write a co-authored paper, researchers must pay in advance all these costs, so it could be expected that researchers tend to prefer working with people whom they have worked with previously and with whom they have already established some norms of cooperation. Moreover, partnering researchers who have already established bond relationships have developed jointly held resources that make the team to cultivate some inertia that makes them more productive. Hence, we expect

Hypothesis 2. Researchers who develop stronger direct relationships will publish more

As noted above, in addition to relationship dimensions, it is also important to consider network structure. One perspective is 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 (Coleman, 1988). In this view, density serves as a proxy for the continual cooperation necessary to succeed in innovation efforts. Yet, others argue (Burt, 1992) that, even if such knowledge sharing occurs, this information will become redundant after some time. Instead, actors embedded in sparsely connected networks will have knowledge brokerage opportunities that they can leverage to construct an efficient, information-rich network, where the redundancy between partners is minimized (Burt, 1992). The notion of structural holes has been proposed to measure the extent to which an individual's position in the network confers the greatest access to novel information and good ideas (Burt, 2004). Empirical studies on the effect of dense and sparse networks are diverse (Rodan and Galunic, 2004). For example, Burt (1992, 2004), Hargadon and Sutton (1997) and Hargadon (2002) found that structural holes facilitate the development of innovative products. Other research, such as Uzzi (1997), Hansen (1999) and Obstfeld (2005), suggests the importance of dense networks in innovation to transfer tacit knowledge.

To reflect on how these perspectives relate to academia, one can consider how knowledge is generated and shared among academic researchers. In academia, there are established contexts and tools to share knowledge. Researchers publish and share the results of their work, as they do not necessarily expect to have intellectual property rights or to generate profits based on it. Moreover, there are congresses, workshops and meetings that contribute to sharing knowledge and enhance wide access to novel information. Yet, the exchange also tends to happen inside relatively closed knit networks within the same domain (Lee and Bozeman, 2005). This process contrasts with the context of industrial innovation, where knowledge sharing is likely to be more complex and less frequent to protect firm secrecy (Gilsing et al., 2008). Given these contexts, one might posit that the benefits of density are less relevant for academic researchers than for people engaged in industrial innovation, where the development of trust and cooperation norms might be quite significant. But the ability to access novel information and good ideas that structural holes brings, appears to be all too relevant. Thus we hypothesize:

Hypothesis 3. Researchers that become embedded in networks richer in structural holes will publish more.

Also related to the structural dimension of the network, it has been found that position in the network might affect the opportunities and constraints of an actor (Hanneman, 2001; Cross and Cumming, 2004). This might be particularly relevant in an academic context 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 important to assess what Wasserman and Faust (1994) characterize as a 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 4. Scientists who become more exposed to other central scientists will publish more.

Finally, we explore the cognitive dimension by looking at knowledge clustering, in particular at the degree of collaboration among scientists within the same discipline, versus collaboration of scientists from different disciplines. Because scientific research is becoming more complex, we have been seeing an increasing division of labor in the profession (Arora and Gambardella 1998). At the same time, collaboration of scientists from different disciplines becomes important to have access to the set of assets and particular skills that allow tackling some of the potentially most interesting problems (Katz and Martin, 1997). Thus, we expect scientists that tend to collaborate with researchers from different disciplines enjoy significant benefits, such that:

Hypothesis 5. Scientists that collaborate more often with researchers from different fields of knowledge will publish more.

3. Method

Our main proposition is that to better understand the impact of network embeddedness on research output it is necessary to analyze the competing effects of all relevant network variables. The components of network embeddedness 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 set of hypotheses 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). Fig. 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 1990s.

The data on publications includes information on:

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

In addition, we had access to personal information on 14,328 researchers, in all fields of knowledge, who have been part of the Mexican National System of Researchers[2] (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[3]

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 and Veloso, 2007), the analysis in this paper is restricted to a sample of 1704 researchers included in the area of Exact Sciences[4] that have been part of SNI at least one year between 1991 and 2002.

We chose this area of knowledge because it is one of the most productive areas in Mexico in terms of number of publications in ISI (Gonzalez-Brambila and Veloso, 2007). Moreover, most of the researchers in this area of knowledge are working in academia and are members of SNI (De la Peña, 2003). Therefore, we have a relatively complete Mexican network and, because they are part of SNI, most researchers have the same incentive to publish, independently of their institutional affiliation. These researchers are also highly concentrated in a few select institutions (De la Peña, 2003), and they tend not to move, a general characteristic of the system in Mexico. This is important because changes of institution would be driven by past success as well as the expectation of increasing productivity, and new location changes the environment to which researchers are exposed, affecting their collaboration networks. This could lead to reverse causality on the assessment of the impact of particular network variables on researcher productivity. Thus, one could argue that this study is not much affected by this issue.

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. Mexican Researchers that are part of SNI publish more than 90% of the total number of Mexican publications in ISI (Conacyt, 2003). The analysis in this paper does not consider the characteristics of the international networks of these researchers because we do not have access to the network relations outside Mexico. Still, we will control for the number of articles with

[2] The Mexican National System of Researchers (SNI) 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 in the country. The production of all researchers is revised periodically to determine if the compensation will continue or not.

[3] 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).

[4] 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.

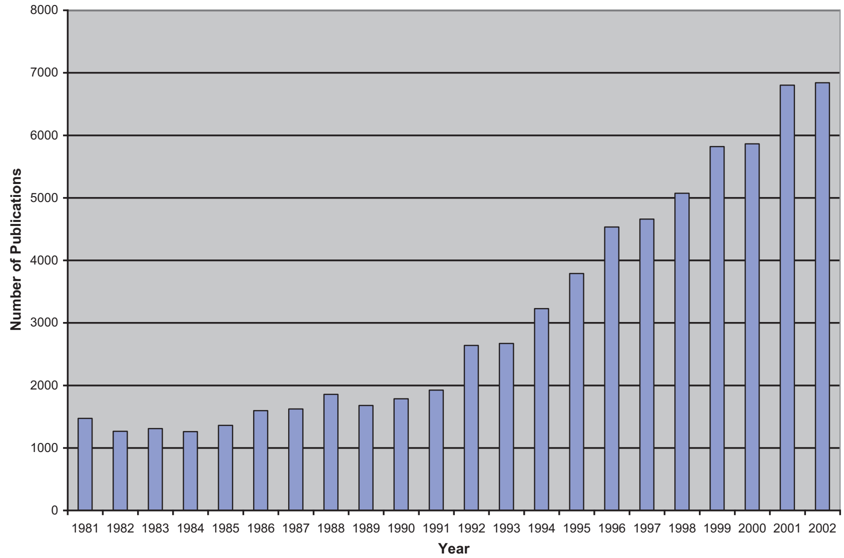

Fig. 1. Evolution of Mexican publications, 1981–2002.

international collaboration, aiming to distinguish heterogeneity across researchers on this dimension.

Since the main purpose of this paper is the analysis of how network embeddedness influences research output, 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" network embeddedness of researchers. Most papers with many authors are the results of large collaborative "big science" projects that conform to a set of procedures and dynamics different from those of smaller groups that we are focusing on. About 97% of the publications have less than 9 authors. So we are including most of the publications. Our analysis is robust to small changes around this definition.

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, such as books, books chapters, proceedings, patents, and non-ISI indexed publications, that are part of the production of researchers and that can be the result of collaboration among researchers, and those are not reflected in our dataset. There are also 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 and we are able to consider a large and rather complete sample of individuals and their peers. Moreover, co-authorship data has the advantage of being objective and replicable for large-sample studies, and perhaps one of the best available data source for studying scientific collaboration (Katz and Martin, 1997).

3.2. Variables

In this paper we measure research output in three ways. First, we use the number of publications that a scientist publishes. Yet, since there might be variations in the quality and impact of the published papers (Lindsey, 1989), we also consider an alternative measure where publications are weighted by the number of cites that each publication receives in the subsequent years. The final measure weights each publication by the number of joint authors.

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 co-authored paper. To study the network of scientists, 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 co-occurrence method was used (Borgatti et al., 2002; Wasserman and Faust, 1994). That is, the co-authorship network is defined on a set of ties among the researchers, where researcher *i* is tied to researcher *j* if and only if researcher *i* and *j* co-authored a publication. This network is often represented as an $N \times N$ adjacency matrix **M**, where N refers to the number of researchers, and cell $\mathbf{M}_{i,j} = 1$ if and only if researchers *i* and *j* co-authored at least one paper; else $M_{i,j} = 0$. The diagonals of this matrix, $\mathbf{M}_{i,i}$, are ignored (or set to missing values), since co-authoring with oneself has no meaning in this context.

The network variables used were:

- Direct ties is the number of unique coauthors during the relevant 3-year period. This is the sum of the rows (or columns, since the network is symmetric) in the adjacency matrix **M**. For example, if researcher A has published with researcher B, C and D in the last 3 years, researcher A has 3 direct ties. Formally, this is referred to as the *degree* of node A in the co-authoriship network.
- Strength of ties draws on a different way of accounting for these co-authorships. The network M only refers to whether the researchers co-authored at least one paper with each of his/her fellow researchers. It is useful, however, to consider how many times a person has written a paper with each of these co-authors. That is, Researcher A may have written one paper with Researcher B but 6 papers with Researcher C. We would say, then, that Researcher A's tie with Researcher C is stronger than A's tie is with Researcher B. Specifically, then, we construct a *valued* network (a network that takes on values beyond just 0 and 1 as indicated in the adjacency matrix **M**). This valued network can also be represented in an $N \times N$ matrix form, **Q**, wherein cells $Q_{i,j}$ = the total number of papers co-authored by researchers *i* and *j* over the 3 year window referred to above (again, the diagonals are ignored). We then construct a "strength of ties" measure of each researcher's involvement with their co-authors as follows. First, we sum the row (or column) values in Q; second, we divide

this row total by the corresponding row total in M. This ratio measures the researcher's propensity to co-author many articles with their colleagues For example, if researcher A has published 2 papers with researcher B, and 6 papers with researcher C (and no papers with anyone else) researcher A has a calculated value of 8/2 = 4 for his/her strength of direct ties.

- Density[5] is often defined across the entire network as the proportion of total ties that exist compared to those that could exist if everyone were tied to everyone else. Operationally, this is usually calculated as the sum of all cell values in M divided by the total number of cells in M (=N × (N − 1)). For our purposes, however, we are concerned about the *local* density that immediately surrounds the particular researcher. That is, we are referring to the extent to which ones co-authors also write papers with each other. To calculate this, we consider a different subgraph for each researcher. That is, for researcher *i* who has co-authored papers with K others in the network M, we extract the K × K subgraph **S** consisting only of those K co-authors. The density for researcher *i* is defined as the density in **S**. For example, if Researcher A had 3 co-authors, B, C and D, and B and C had co-authored with each other but neither had written with D, then the density score for Researcher A would be 1/3 (1 tie: B with C; 2 non-ties between B and D and between C and D). This local density measure indicates the extent of collaboration or cohesion among the co-authors that each researcher is engaged with.
- Structural holes are the separation of different actors who are not connected, the absence of ties between two parts of the network. 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 an ego is tied to people who are in turn tied to each other, thereby creating lots of redundancy in ego's local network.
- Centrality is measured by using the normalized eigenvector proposed by Bonacich (1972). This measure of centrality captures the important feature that an ego's status and power in a network is a function not only of how many alters they are tied to but also how high in centrality (and consequent status and power) each of these alters is. That is, a high value is given to an actor who is connected to many actors who are themselves also well-connected (Borgatti, 1995).
- 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; Tortoriello and Krackhardt, 2010). This index ranges from −1 to +1 and indicates the extent to which the network ties cut across group boundaries. For example, a "−1" would indicate that all network ties occur within groups; a +1 would indicate that all network ties cut across group boundaries; a "0" would indicate that exactly half the ties occur within groups and half the ties cut across groups. Originally, the E–I index was designed to measure the extent to which organizations as a whole were characterized as having network ties that cut across formal organizational boundaries, such as divisions or departments. More recently, though, the index has been used to measure and individual's propensity to network within their "group" or to bridge across groups with their network ties.

The groups, in this case, were "fields of knowledge". That is, the E–I index calculated for each author measures the extent to which he/she tends to co-author with scholars from different fields than their own. We lacked definitive information on each author's primary field of knowledge, especially for those who have not been part of SNI. However, we do know the field in which each author published first. We assumed that this first publication is a proximate indication of the researcher's primary field of knowledge and used this marker to identify the researcher's field.

All these network variables were calculated using UCINET (Borgatti 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:

$P_{it} = F(X_{it-1}, c_i, u_{it})$, *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, density, structural holes, centrality, external-internal index, reputation (description below).c_i: is the individual unobserved effect which is stable across time but not across researchersu_{it}; is the error term

We use the negative binomial fixed effects model proposed by Hausman et al. (1984) because of the panel nature of the data. This is aligned with our objective of analyzing within person variations to understand how the social embeddedness around an individual conditions relevant outcomes, in this case scientific output. The fixed effects model explores the temporal variation of critical variables for the focal individual, thus allowing 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. Simultaneously, this empirical approach also allows a control for any fixed unobserved heterogeneity across the institutional setting of the individual, an important influence of research output and impact. This means that we will be exploring how the current (and recent past) nature of the ego-network characteristics for each individual influences their future output (and impact). To assure the consistency of our approach, 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. Yet, a drawback of the Negative Binomial distribution is that the conclusions may be less precise because the estimated standard errors tend to be larger than in the alternative Poisson model.

As stated before, three different measures of research output were created. The first one (pubs) measures the straight publication counts occurring over a two-year period. We decided to consider publication output over 2 years because publication is an uneven event and many researchers in Mexico do not publish every year in ISI publications. By using this measure we avoid losing observations from having many zeros in the outcomes. The second output measure (cites) adjusts publications for quality by adding the number of cites that publications have received in the subsequent 4 years. This citation window was chosen trying to balance the loss of observation years that result from considering forward citations, against the desire to include as much as possible this measure of quality and impact.[6] On average for our sample, publications receive 70%

[5] Following Obstfeld (2005), we used two measures of structural holes, density and Burt's (1992) measure of structural holes (1-constraint).

[6] This approach is similar to the one used by Gonzalez-Brambila & Veloso (2007) who provide a discussion on the robustness of this approach. An alternative possibility would be to consider the average number of citations per year. The problem with this solution is that more recent publications would have smaller average number of citations, which could bias our results. Alternatively, one could consider the indicator 'expected number of citations' produced by ISI. We decided not to consider this approach because there is not enough evidence that this ISI indicator better reflects the quality and significance of a given publication.

Table 1
Periods.

Period	*T* (output variable)	*t*-1 (network variables)
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

Table 2
Descriptive statistics.

Variable	Mean	Std. Dev.	Min	Max
Pubs	1.7642	3.0050	0	53.0
Cites	6.3630	18.3926	0	506.0
Frac pubs	0.6337	1.0933	0	14.3
Direct ties	6.2745	6.8184	1	148.0
Strength dt	1.4534	0.6599	1	10.0
Density	82.7719	30.1232	0	100.0
Structholes	0.4120	0.2952	0	1.0
Eigenvec	0.3438	4.0750	0	101.9
E-I Index	−0.3538	0.7259	−1	1.0
Sing auth	0.1694	0.6775	0	17.0
Int art	0.7580	2.0038	0	44.0
Past pubs	2.5619	4.6152	0	69.0
Past cites	8.2052	23.8579	0	645.0
Past frac pubs	0.7092	1.4147	0	28.1
Observations 15336				
n = 1704				
T = 9				

of cites in this window, with the remaining 30% received in the following 10 years. The final measure, (frac pubs) adjusts the count of publications by summing the inverse of the number of coauthors of each publication.

To calculate the relevant network variables for the focal researchers, we do not look at the contemporary characteristics of the network, but rather at their immediate past. This means that we will be assessing how these recent past characteristics explain future scientific performance. This approach avoids problems of simultaneity between outcome and explanatory variables in the regression, a particular salient issue in network analysis. Yet, we are still considering the network variables at any given moment and for a limited time, rather than the network structure based on the entire stock of past relations. In particular, the network variables for researcher i at time t-1, where t corresponds to the 2 year publication window explained above, were obtained from the adjacent matrixes[7] considering information on the publications during the previous 3 years. For example, if output for time t corresponds to the years 1999–2000, the network variables used as covariates will be calculated using the publications during the period 1996–1999. The decision to use the previous 3 years comes from the need 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.

Thus, nine periods were obtained as is showed in Table 1.

In an effort to isolate the effects of network structures in productivity, a number of additional control variables were considered in the analysis. Considering the large growth in the number of publications that occurred over the period of analysis (Fig. 1), time dummies were included to capture time trends. However, it is important to note that including time dummies prevents us from considering researcher age or time since PhD in the regressions as a control because of 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 included in the panel structure, it is possible that changes in reputation also 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 favors those researchers who already have a 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; 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 ensure that greater levels of output would be indeed resulting from the social embeddedness the researcher is building, and are not confounded effects with the reputation developed by the researchers.

Two different measures were used to control for reputation: (1) the number of single-authored papers that the researcher has published in the past three years, and (2) the international visibility by counting the number of articles with a foreign address in the same three-year window. We believe that these two indicators are reasonable measures of reputation since most universities used at least one of them to give raises and promotions (Tien and Blackburn, 1996; McFadyen and Cannella, 2004)

Finally, we also considered the variables past publications,[8] or past cites because including a lagged dependent variable helps to control for unobserved individual time variant variables and for other potentially important, but possible omitted predictors (Greene, 2000).

4. Results

Table 2 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. The average number of coauthors is 3.35 and the average number of institutions is 2.5. Each of these researchers 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. These figures are similar to those in Singh (2007), a study that is focused on Biotechnology and Applied Biology, where the average number of direct ties is 4.80 and the strength of those ties is 1.20. However, these numbers contrast with the study of McFadyen and Cannella (2004) focusing on biomedical scientists, where the number of relations is substantially larger (46.5), while the strength of those relations is lower (1.33). In our study, the average density is quite high (82.8), while the mean of the structural holes variable is 0.41. However, given that the sample is large, it is possible that the structural holes' variable tends to be higher due to the size of the network (Burt, 1998). The mean of centrality is 0.34, which may seem too low.

[7] 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.

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

Table 3
Correlations.

Variable	Pubs	Cites	Frac pubs	Direct ties	Strength dt	Density	Structholes	Eigenvec	E–I Index	Sing auth	Int art	Past pubs	Past cites
Pubs	1												
Cites	0.6952	1											
Frac pubs	0.8785	0.5849	1										
Direct ties	0.5469	0.4247	0.3291	1									
Strength dt	0.2619	0.1824	0.2081	0.1831	1								
Density	−0.4010	−0.2675	−0.3659	−0.5448	−0.0978	1							
Structholes	0.3633	0.2571	0.2546	0.6428	0.0825	−0.7266	1						
Eigenvec	0.2647	0.3519	0.1767	0.3260	0.1385	−0.0946	0.0957	1					
E–I Index	0.0051	−0.0697	−0.0286	0.0681	−0.0226	−0.0471	0.0598	−0.0539	1				
Sing auth	0.2017	0.1495	0.3964	0.0119	0.0136	−0.1433	0.0366	0.0255	−0.0063	1			
Int art	0.4779	0.5592	0.3895	0.5659	0.2891	−0.3732	0.3405	0.5036	−0.0664	0.1259	1		
Past pubs	0.6547	0.5229	0.5577	0.7656	0.4481	−0.5735	0.5007	0.3661	0.0198	0.2924	0.6740	1	
Past cites	0.5458	0.6520	0.4659	0.5347	0.3147	−0.3507	0.3203	0.4617	−0.0759	0.2000	0.7460	0.6946	1

However, it is important to note that this is the result of a wide dispersion, with a few people whose eigenvector is quite large (102) and a lot that are not central at all. The mean of the external-internal index is negative, which suggests that scientists tend to collaborate more with peers within the same area of knowledge. This was to be expected because of the fields that are included in this area of knowledge.

Table 3 shows the correlation between our critical variables. As can be seen, there are high correlations among some variables. In particular, there is a high correlation between the number of direct ties and past publications. One could conceive that such high correlation among some of the variables used in the regression could cause a multicollinearity problem. Yet, it is important to remember that, even if that were the case, unbiased estimators would still be produced, and this problem would only increase the variances of the collinear variables (Kennedy, 2003). Therefore, finding statistical significance not only means that we have valid results, but also that their true significance levels are probably understated.[9]

Considering that our use of panel data could be associated with autocorrelation among variables, Wooldridge (2002, pp 282–283) tests were performed to be certain that our analyses are not affected by this potential issue. The results of the tests reject the possibility of autocorrelation.

Table 4 reports the regressions results of the fixed effects negative binomial model[10] using the straight count of publications as dependent variable.

The results confirm that the relative contribution of each variable is moderated when other network variables are introduced. The positive effect of direct ties, that would have confirmed hypothesis 1, disappears when other variables are included. The negative effect of the strength of direct ties, that would have rejected hypothesis 2, also vanishes in the complete model. The negative effect of density is lightly moderated in the broad model, so that we could think that hypothesis 3 is confirmed. However, the effect of structural holes also disappears, so that hypothesis 3 cannot be confirmed. The positive effect of centrality and cognitive dimension, that confirms hypotheses 4 and 5, remains and does not change marginally.

In all specifications models, single-authored papers is the only control variable that is significant. As expected, it positively affects the future productivity of researchers. Surprisingly, international articles is not significant.

Table 5 reports the results of the fixed effects negative binomial model now using the number of cites in the next four years of the publication. Given that the last year of information is 2002; only periods 1–7 were included.

As in Table 4, most of the impacts of the network variables are moderated in the complete models (7 and 8).

The results show that when adjusting for quality, the results change a lot compared to those when the number of publication is used. In this case, the relational dimension (direct ties and strength of direct ties) affects positively the quality of research outputs,

[9] Different regressions were run dropping past publications as a control variable. In particular we were concerned with using past publications as a control variable because of multicollinearity problems with direct ties. The results were not different from the regressions where past publications and direct ties were included. So, we decided to leave both variables in the regressions shown in this paper. Given that including a lagged dependent variable reduces omitted variable problems (Greene, 2000) we report the results considering this control variable.

[10] To addressed the concern that the conditional fixed effects negative binomial does not account for over-dispersion nor does control for any fixed unobserved heterogeneity across individuals (Allison and Waterman, 2002), we run a new model using the "xtqmlp" procedure (available for download at http://scripts.mit.edu/~pazoulay/doc/xtqmlp.ado). The results obtained with this specification were similar in sign and significance to those obtained with a negative binomial fixed effects estimation. The results could be shown upon request. For comparability with previous research on scientific productivity that commonly employed negative binomial regressions, we decide to show the results obtained by using negative binomial fixed effects models.

Table 4
Regressions results for publications.

	(1) pubs	***m*** pubs	(3) pubs	(4) pubs	(5) pubs	(5) pubs	(7) pubs	(8) pubs
Direct ties	0.0068**						0.0037	0.0048
	(0.0022)						(0.0025)	(0.0025)
Strengthdt		−0.0358*					−0.0160	−0.0214
		(0.0175)					(0.0135)	(0.0134)
Density			−0.0050***				−0.0019***	
			(0.0004)				(0.0004)	
Structholes				0.120*				0.0695
				(0.0489)				(0.0516)
Eigenvec					0.0058**		0.0061***	0.0059**
					(0.0013)		(0.0018)	(0.0018)
E–l Index						0.0776**	0.0685**	0.0735**
						(0.0249)	(0.0252)	(0.0251)
Past pubs	−0.0038	−0.0036	−0.0065	0.0004	0.0013	0.0012	−0.0056	−0.0035
	(0.0035)	(0.0029)	(0.0033)	(0.0029)	(0.0028)	(0.0028)	(0.0038)	(0.0038)
Single auth	0.0353**	0.0270*	0.0897***	0.0298*	0.0323*	0.0284*	0.0351**	0.0356**
	(0.0128)	(0.0125)	(0.0119)	(0.0126)	(0.0126)	(0.0126)	(0.0130)	(0.0129)
Int art	−0.0000	0.0021	0.0118*	0.0018	0.0004	0.0030	0.0005	0.0004
	(0.0052)	(0.0051)	(0.0056)	(0.0052)	(0.0051)	(0.0051)	(0.0051)	(0.0051)
Cons	2.002***	2.052***	1.577***	1.951***	2.021***	2.045***	2.198***	2.049***
	(0.0700)	(0.0726)	(0.0539)	(0.0736)	(0.0696)	(0.0707)	(0.0787)	(0.0800)

Standard error in parentheses
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.
The regression models include fixed effects by indivicual and time.

Table 5
Regressions results for cites.

	(1) Cites	**P)** Cites	(3) Cites	**1 × 1** Cites	(5) Cites	(5) Cites	(7) Cites	(8) Cites
Direct ties	0.0250***						0.0175***	0.0201***
	(0.0037)						(0.0043)	(0.0044)
Strengthdt		0.0800**					0.104***	0.101***
		(0.0271)					(0.0273)	(0.0274)
Density			−0.0099***				−0.0039***	
			(0.0006)				(0.0007)	
Structholes				0.410***				0.269**
				(0.0805)				(0.0911)
Eigenvec					0.0023		0.0037	0.0027
					(0.0035)		(0.0034)	(0.0034)
E–l Index						0.0029	−0.0252	−0.0214
						(0.0365)	(0.0373)	(0.0372)
Past cites	0.0018*	0.0021**	0.0025**	0.0022**	0.0022**	0.0023**	0.0013	0.0014
	(0.0008)	(0.0008)	(0.0008)	(0.0008)	(0.0003)	(0.0007)	(0.0008)	(0.0008)
Single auth	0.0875***	0.0805***	0.130***	0.0809***	0.0816***	0.0798***	0.0860***	0.0895***
	(0.0209)	(0.0215)	(0.0170)	(0.0212)	(0.0216)	(0.0215)	(0.0210)	(0.0208)
Int art	0.0008	0.0164	0.0049	0.0141	0.0185	0.0202	−0.0087	−0.0059
	(0.0113)	(0.0109)	(0.0111)	(0.0110)	(0.0112)	(0.0109)	(0.0117)	(0.0117)
Cons	−0.398***	−0.346***	0.0569	−0.443***	−0.219***	−0.225***	−0.293***	−0.652***
	(0.0588)	(0.0661)	(0.0539)	(0.0676)	(0.0526)	(0.0526)	(0.0866)	(0.0845)

Standard error in parentheses
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.
The regression models include fixed effects by individual and time.

confirming hypotheses 1 and 2. In terms of the structural dimension, structural holes has a strong positive effect, confirming hypothesis 3. These 3 results together are consistent with those obtained by Singh (2007) that uses a worldwide sample of scientists in Biotechnology and Applied Microbiology. Density, as well as in the case of publications, affects negatively. The positional dimension, as well as the cognitive dimension, does not have an impact on the future quality.

As in the case of publications, the single authored papers control variable is significant in 8 models. Reputation, measured as the accumulated number of cites per researcher is highly significant in the models when each network variable is included in the model. When all network variables are included in the model, it is significant at 10%. These results reinforce the Mathew effect in science (Merton, 1968).

Finally, Table 6 shows the results of a fixed effects ordinary least square model, where the fractional number of publications times 10[11] is the dependent variable. The results, in terms of sign and significance, are very similar to those obtained for the straight number of publications, although they are less significant.

Finally, we also tested 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, for both measures of productivity, straight count of publications and number of cites in the next four years, we find no evidence of diminishing returns in either the number of direct ties or the strength of those ties.[12]

5. Discussion and conclusions

This study looks at how network embeddedness around a focal researcher conditions his or her output and impact. We address this issue using a unique dataset of Mexican scientific publications in the Exact Sciences area. First, we control for unobserved heterogeneity as well as for a set of other critical observables that can condition performance, including past reputation. Thus, we explore how cross time changes in the network structure of the individual correlate with performance. Second, we use objective measures of network embeddedness, thus not sensitive to the subjective bias that comes with interviews. Third, we are able to analyze a very comprehensive network. Fourth, we use different measures of productivity. Finally, we incorporate most relevant network metrics simultaneously in the analysis.

The results offer relevant contributions to our understanding of network theory and practice, as well as to the specific study of the dynamics of scientific collaboration. First, in terms of scientific collaboration, our study shows that the network dynamics behind the generation of quality output contrasts dramatically with that of quantity. We find that the relational dimension of scientist matters for quality, but not to output, while cognitive dimensions has the opposite effect, helping output, while being indifferent toward impact. Yet, the most significant result is probably the establishment of some degree of prevalence for brokerage over cohesion in the role that the structural dimension of the network plays in performance. As noted in the introduction and theory sections, this is a longstanding tension that has not been resolved by the literature. We believe our results advance our understanding of these tensions.

The structural dimension of the network is the only area where we find some degree of convergence between quantity and quality of output, even controlling for unobserved heterogeneity across individuals. First, we find a negative impact of density for both

Table 6
Regressions results for fractional number of publications.

	(1) (Frac pubs × 10)	(2) (Frac pubs × 10)	(3) (Frac pubs × 10)	(4) (Frac pubs × 10)	(5) (Frac pubs × 10)	(6) (Frac pubs × 10)	(7) (Frac pubs × 10)	(8) (Frac pubs × 10)
Direct ties	0.0053						0.0032	0.0096
	(0.0308)						(0.0323)	(0.0337)
Strength dt		−1.646*					−1.779	−1.762
		(0.224)					(0.228)	(0.228)
Density			−0.0113***				−0.0165***	
			(0.0032)				(0.0049)	
Structholes				−1.220				−1.949
				(0.547)				(0.598)
Eigenvec					0.0130*		0.0212*	0.0199*
					(0.0354)		(0.0353)	(0.0354)
E–I Index						0.0722*	0.135*	0.132*
						(0.276)	(0.275)	(0.275)
Frac Past pub	2.021***	2.220***	3.246***	2.080***	2.026***	2.026***	2.357***	2.299***
	(0.115)	(0.111)	(0.079)	(0.111)	(0.108)	(0.109)	(0.122)	(0.120)
Single auth	−1.648***	−1.843***	−1.154***	−1.710***	−1.651***	−1.655***	−1.966***	−1.923***
	(0.227)	(0.223)	(0.147)	(0.224)	(0.223)	(0.223)	(0.229)	(0.229)
Int art	−0.215**	−0.151*	−0.096	−0.196**	−0.214**	−0.209**	−0.130	−0.136
	(0.0773)	(0.0722)	(0.0512)	(0.0724)	(0.0728)	(0.0722)	(0.0785)	(0.0785)
cons	8.507***	10.50***	3.853***	9.040***	8.538***	8.547***	9.542***	11.46***
	(0.374)	(0.441)	(0.349)	(0.419)	(0.352)	(0.358)	(0.594)	(0.530)

Standard error in parentheses
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.
The regression models include fixed effects by indivicual and time.

[11] To avoid the zero truncation, the variable was multiplied by 10.
[12] Results of these regressions are shown upon request.

measures of output. Furthermore, there is a positive impact for structural holes in the number of cites, often considered the most impactful output. This implies that non-redundant information that comes from brokerage is more beneficial than the coordination that is obtained in dense networks. This positive impact of brokerage on cites is consistent with a variety of prior studies (Burt, 1992, 2004; Singh, 2007), which suggest that structural holes bring non-redundant information that benefits output. In terms of the structural dimension, we also find that output is affected positively by centrality, lending some support to the idea that position in the network affects the opportunities of an actor (Hanneman, 2001). However, centrality does not necessarily convert in impact, leading us to conclude that such benefits have some limitation, especially in terms of quality.

When considering the relational dimension between scientists, the analysis shows that researchers who invest resources in having many and frequent ties are able to generate greater impact, albeit without a significant gain in the number of publications. This suggests that the breadth and depth of personal relationships scientists develop with each other through the publishing effort might indeed bring a diversity of good ideas (Burt, 2004), which results in higher quality publications. On the contrary, collaboration across discipline boundaries, a cognitive dimension, is associated only with increased output. This could be the result of a growing division of academic labor (Adams et al., 2005), which may require interdisciplinary and multidisciplinary collaborations. Bridging across communities gives researchers ideas and resources for new papers, but these do not necessarily capture sufficient attention from the different communities and thus do not receive as many cites.

Our results also bring interesting new perspectives to established notion that collaboration in research is important and should be promoted (Katz and Martin, 1997). Consistent with most existing work, we also conclude that the social embeddedness that results from scientific collaboration matters. Yet, our research advances beyond existing work by providing particular insights on how initiatives to foster research collaboration can take into account the relational, structural and cognitive dimensions of social capital. In particular, our results suggest that relational efforts (greater number of coauthors, stronger bonds) matter for impact, albeit they will not be helpful in terms of sheer output. This finding also provides a novel and more precise perspective on the long-established notion that high level-impact science needs some degree of critical mass (Stephan, 1996; Gonzalez-Brambila and Veloso, 2007). In the future, when considering actions toward developing critical mass, rather than focusing on the amount of active scientists, which is the more typical perspective, decision makers may want to consider instead what supports the establishment and deepening of the relational dimension noted here. Similarly, the research also highlights an interesting tradeoff on the role of interdisciplinary work (Katz and Martin, 1997) and network centrality. While bridging disciplines and moving toward a central position contributes to output, the result seems to have less of an impact on the profession. Overall, our work shows there is a clear trade-off between quality and quantity in research productivity when investing and leveraging network assets.

Our work also makes a contribution to our understanding of network theory and practice. Our results support the notion that one should adopt a contingency perspective when studying the impact of network embeddedness on performance (Phelps et al., 2012). On one hand, the relative contribution of each variable is clearly moderated by other network variables (Singh, 2007; McFadyen et al., 2009). In fact, the impact of some of the variables disappears altogether when other critical variables are included. This moderation appears to be particularly salient in social embeddedness, with explanatory variables for tie strength or depth having statistical significance when considered individually, but losing this significance when including a broader set of network covariates in the statistical analysis. On the other hand, the various dimensions of network embeddedness have different effects depending on the indicator of performance that is used (Gonzalez-Brambila and Veloso, 2007; He et al., 2009). The relational dimension of network embeddeness appears to be important and stable when considering creative or insightful output, which we equated to citations measurements. But it does not matter for run-of-the-mill results, where structural and cognitive variables play the more significant role. Perhaps not surprisingly, structural variables play a more consistent role across performance variables.

Our results also offer a different light on the validity of the arguments on the roles that network embeddedness play in individual performance. The various inputs of network embeddedness influence the quantity and quality of outputs in a diverse way. These results help explain the range of results of previous empirical studies in network embeddedness, which typically do not consider a range of output variables. They also provide specific implications for future work. By contrasting how the various network embeddedness dimensions condition more creative or impactful output vs. routine results, future research could help validate our results and develop a more stable contingency pattern in terms of how the nature of the network impacts research output and impact.

As expected, controlling for single authored papers is important. Although there has been an increase in collaboration and academic team size over the last years (Adams et al., 2005), this study shows that the most productive and collaborative researchers are also those more likely to publish single authored papers. Surprisingly, international collaboration seems not to affect significantly future output and impact, as other studies have found (Singh, 2007; Lee and Bozeman, 2005; He et al., 2009). One possibility is that researchers who are publishing in ISI journals, a criteria for observing output in our study, are also those with consistent international collaborations, such that we do not have much cross time variation to explain performance changes; the gaps between faculty are persistent and thus absorbed by the fixed effects. However, because our study only controls for international visibility and we do not have the characteristics of any international links, it is hard to provide a good interpretation for this result.

This study has also limitations that may condition our results in certain ways. First, an aspect that could be confounded with the effects of changes in networks variables on the productivity of individuals is individual affiliation. In countries with researcher mobility, changes of institution would be driven by past success, as well as the expectation of increasing productivity. But when they move, the new location allows dramatic changes in the environment to which they are exposed, affecting their collaboration networks. This could lead to reverse causality on the assessment 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. But, on the downside, it can also be considered that our sample does not capture how significant changes in the network embeddedness of individuals conditions performance. Second, co-publication is not an exhaustive measure of collaboration; there are more products of collaboration than joint publications in ISI (Katz and Martin, 1997). 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).

Not less important is the fact that network embeddedness is very difficult to measure because it has many and varied components, some of them intangible. This study has not considered all the relationships that someone could have and that could affect, to some degree, their future output and impact. 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 network embeddedness 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 productivity.

Another limitation of the study is that it does not include the costs associated with network embeddedness. The benefits related with future productivity should be weighed against the costs of collaboration itself, and the costs of managing and maintaining the different dimensions of social capital.

Finally, it is acknowledged that network embeddedness could be very different 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 apply equally to all fields of knowledge. The other relevant consideration is that, although it could seem that all Mexican researchers have the same incentive to publish, since the SNI system is open to every researcher, in Mexico, as in the rest of the world, the productivity of researchers is highly skewed. A future line of research could analyze specifically the differences in the network embeddedness of highly productive researchers versus those who are not, including differences in areas of knowledge, and exploring how international collaboration could affect future output and impact.

In summary, while the results of this study should not be considered completely conclusive, given the limitations stated above, we believe that the work offer some relevant insights. The most significant result is the establishment of some prevalence for brokerage over cohesion in the role that the structural dimension of the network plays in performance, a longstanding tension in the literature. Second, the study highlights how the characteristics of network embeddedness play different roles depending on the measure of productivity. Paper outputs seem to be most helped by non-dense networks, being central and collaborating with researchers from different areas of knowledge. However high impact can be achieved through more and frequent collaborations, and environments rich in structural holes. Third, this study shows that when different dimensions of network embeddedness are included, the positive effect of most network variables diminishes or vanishes, suggesting that it is possible that other studies have tended to overestimate the isolated effects of the different dimensions of social capital. The results of this study could help policy makers and university administrators to designate resources to stimulate certain characteristics of network embeddedness instead of promoting all kinds of collaboration.

Acknowledgments

Claudia Gonzalez-Brambila would like to thank the Instituto Tecnológico Autónomo de México (ITAM) and the Asociación Mexicana de Cultura AC for their support of this work. Francisco Veloso would like to acknowledge the support of the ZON Chair in Innovation and Entrepreneurship at Católica-Lisbon.

References

Adams, J.D., Black, G.C., Clemmons, J.R., Stephan, P.E., 2005. Scientific teams and institutional collaborations: evidence from U.S. Universities, 1981–1999. Research Policy 34 (3), 259–285.

Ahuja, G., 2000. Collaboration networks, structural holes, and innovation: a longitudinal study. Administrative Science Quarterly 45 (3), 425–455.

Allison, P.D., Waterman, R.P., 2002. Fixed-effects negative binomial regressions models. Sociological Methodology 32 (1), 247–265.

Arora, A., Gambardella, A., 1998. The Impact of NSF Support for Basic Research in Economics. Working paper.

Arora, A., David, P., Gambardella, A., 1998. Reputation and competence in publicly funded science. Annales d'Economie et de Statistiques 49/50, 163–196.

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.

Beaver, D.R., Rosen, 1978. Studies in Scientific Collaboration. Part I. The Professional Origins of Scientific Co-authorship. Scientometrics 1 (1), 65–84.

Bonacich, P.B., 1972. Technique for analyzing overlapping memberships. In: Costner, H. (Ed.), Sociological Methodology. Jossey-Bass, San Francisco, pp. 176–185.

Borgatti, S.P., 1995. Centrality and AIDS. Connections 18 (1), 112–115.

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

Burt, R.S., 1992. Structural holes: the social structure of competition. Harvard University Press, Cambridge.

Burt, R.S.,1998. The network structure of social capital. In: Paper presented at the Social Network and Social Capital Conference. Duke University, Durham, NC.

Burt, R.S., 2001. Structural holes versus network closure as social capital. In: Nan Lin, Cook, K.S., Aldine de GruyterHiraki, R.S.B. (Eds.), Social Capital: Theory and Research.

Burt, R.S., 2004. Structural holes and good ideas. American Journal of Sociology 110 (September), 349–399.

Coleman, J.S., 1988. Social capital in the creation of human capital. American Journal of Sociology 94, 95–120.

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

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

De la Peña, J.A., 2003. La Ciencia en México. AMC.

Fleming, L., Mingo, S., Chen, D., 2007. Collaborative brokerage, generative creativity, and creative success. Administrative Science Quarterly 52 (3), 443–475.

Freeman, C., 1982. The Economics of Industrial Innovation. MIT Press, Cambridge, MA.

Gilsing, V., Nooteboom, B., Vanhaverbeke, W., Duysters, G., Oord, A., 2008. Network embeddedness and the exploration of novel technologies: technological distance, betweens centrality and density. Research Policy 37, 1717–1731.

Glanzel, W., Schubert, A., 2001. Double Effort=double impact? a critical view at international co-authorship in chemistry. Scientometrics 50 (2), 199–214.

Gonzalez-Brambila, C., Veloso, F., 2007. The determinants of research output and impact: a study of mexican researchers. Research Policy 36 (September), 1035–1051.

Granovetter, M., 1973. The strength of weak ties. American Journal of Sociology 78 (6), 1360–1380.

Granovetter, M., 1992. Problems of explanation in economic sociology. In: Nohria, N., Eccles, R.G. (Eds.), Networks and Organizations. Harvard Business School Press, Boston, MA, pp. 25–56.

Greene, W.H., 2000. Econometric Analysis, 4th ed. Prentice Hall, Upper Saddle River, NJ.

Gulati, R., 1995. Social structure and alliance formation: a longitudinal analysis. Administrative Science Quarterly 43 (4), 781–814.

Hanneman, R., 2001. Introduction to Social Network Methods. Department of Sociology. University of California, Riverside.

Hargadon, A., Sutton, R.I., 1997. Technology brokering in a product development firm. Administrative Science Quarterly 42 (4), 716–749.

Hargadon, A., 2002. Brokering knowledge: learning and innovation. In: Sutton, R.I., Staw, B.M. (Eds.), Research in Organizational Behavior. Elsevier/JAI, New York, pp. 41–85, 24.

Hausman, J.A., 1978. Specification tests in econometrics. Econometrica 46 (Nov), 1251–1271.

Hausman, J., Hall, B., Griliches, Z., 1984. Econometric models for count data with an application to the patents-R&D relationship. Econometrica 52 (July), 909–938.

Hansen, M.T., 1999. The search-transfer problem: the role of weak ties in sharing knowledge across organization subunits. Administrative Science Quarterly 44 (1), 82–111.

He, Z.L., Geng, X.S., Campbell-Hunt, C., 2009. Research collaboration and research output: a longitudinal study of 65 biomedical scientists in a New Zealand university. Research Policy 38, 306–317.

Inkpen, A.C., 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.

Katz, J.S., 1994. Geographical proximity of scientific collaboration. Scientometrics 31 (1), 31–43.

Katz, J.S., Hicks, D., 1997. How much is a collaboration worth? a calibrated bibliometric model. Scientometrics 40 (3), 541–554.

Katz, J.S., Martin, B.R., 1997. What is research collaboration? Research Policy 26, 1–18.

Kennedy, P., 2003. A Guide to Econometrics, 5th ed. The MIT Press, Cambridge, MA.

Krackhardt, D., Stern, R.N., 1988. Informal networks and organizational crises: an experimental simulation. Social Psychology Quarterly 5 (2), 123–140.

Larson, A., 1992. Network dyads in entrepreneurial settings: a study of the governance of exchange processes. Administrative Science Quarterly 37 (1), 76–104.

Lee, S., Bozeman, B., 2005. The impact of research collaboration on scientific productivity. Social Studies of Science 35 (5), 673–702.

Levin, S., Stephan, P., 1991. Research productivity over the life cycle: evidence for academic scientists. The American Economic Review 81 (March), 114–132.

Lindsey, D., 1989. Using citation counts as a measure of quality in science: measure what's measurable rather than what's valid. Scientometrics 15 (3–4), 189–203.

Lotka, A.J., 1926. The frequency distribution of scientific productivity. Journal of the Washington Academy of Sciences 16 (12), 23–317.

McFadyen, M.A., 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.

McFadyen, M.A., Semadeni, M., Cannella, A.A., 2009. Value of strong ties to disconnected others: examining knowledge creation in biomedicine. Organization Science 20 (3), 552–564.

Melin, G., 1999. Impact of national size on research collaboration: a comparison between northern European and American universities. Scientometrics 46 (1), 161–170.

Melin, G., Persson, O., 1996. Studying research collaborations using co-authorship. Scientometrics 36 (3), 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., Ghoshal, S., 1998. Social capital, intellectual capital, and the organizational advantage. The Academy of Management Review 23 (2), 242–266.

Narin, F., Stevens, K., Whitlow, E., 1991. Scientific cooperation in Europe and the citation of multidomestically authored papers. Scientometrics 21 (3), 313–323.

Obstfeld, D., 2005. Social networks, the tertius iungens orientation, and involvement in innovation. Administrative Science Quarterly 50 (1), 100–130.

Phelps, C., Heidl, R.A., Wadhwa, A., 2012. Knowledge, networks, and knowledge networks: a review and research agenda. Journal of Management 38, 1115–1166.

Price, D., Beaver, D., 1966. Collaborating in an invisible college. American Psychologist 21, 1011–1018.

Reagans, R., McEvily, B., 2003. Network structure and knowledge transfer: the effects of cohesion and range. Administrative Science Quarterly 48 (2), 240–267.

Reagans, R., McEvily, B., 2008. Contradictory or compatible? Reconsidering the "tradeoff" between brokerage and closure on knowledge sharing. Network Strategy – Advances in Strategic Management 25, 275–313.

Rodan, S., Galunic, C., 2004. More than network structure: how knowledge heterogeneity influences managerial performance and innovativeness. Strategic Management Journal 25 (6), 541–562.

Rowley, T., Behrens, D., Krackhardt, D., 2000. Redundant governance structures: an analysis of structural and relational embeddedness in the steel and semiconductor industries. Strategic Management Journal 21, 369-286.

Singh, J., 2007. External Collaboration, Social Networks and Knowledge Creation: Evidence from Scientific Publications. INSEAD, Mimeo.

Stephan, P., 1996. The economics of science. Journal of Economic Literature XXXIV, 1199–1235.

Stephan, P., Levin, S., 1997. The critical importance of careers in collaborative scientific research. Revue d'Economie Industrielle 79 (1), 45–61.

Tien, F.F., Blackburn, R.T., 1996. Faculty rank system, research motivation, and faculty research productivity: measure refinement and theory testing. The Journal of Higher Education 67 (1), 2–22.

Tortoriello, M., Krackhardt, D., 2010. Activating cross-boundary knowledge: the role of Simmelian ties in the generation of innovations. Academy of Management Journal 53, 167–181.

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.

Uzzi, B., 1997. Social structure and competition in interfirm networks: the paradox of embeddedness. Administrative Science Quarterly 42 (2), 35–67.

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

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

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. Science 316 (5827), 1036–1039.

Zucker, L.G. and Darby, M.R., Movement of star scientists and engineers and high-tech firm entry, National Bureau of Economic Research, Working Paper No. 12172, April, revised October, 2006.

Zuckerman, H., 1967. Nobel laureates in science: patterns of productivity, collaboration and authorship. American Sociological Review 32 (3), 391–403.