Network effects of Collaboration on Scientific Productivity and Impact: A Multidisciplinary Author Level Analysis

4th June 2019

1 Introduction

More than 20 years ago a change in the means of production of scientific and technical knowledge was identified, with research becoming more interdisciplinary and increasingly conducted in a growing body of domestic and international networks (D. M. Hicks & Katz, 1996). This change in the collaborative behaviour of researchers has not changed or diminished. On the contrary, macro level patterns suggest that the density of knowledge networks that reflect the connectivity between authors has actually been increasing (Leydesdorff & Wagner, 2008).

In addition to the observation that the collaborative behaviour of researchers has increased in the process of knowledge production, it has also been shown that there is a positive correlation between scientific performance and collaborative research endeavours (Wuchty, Jones & Uzzi, 2007).

Aside from the interest of researchers in finding evidence of positive gains of knowledge sharing, there are also some policy measures specifically designed to increase the interaction of the science system that might be justified by the assumption that more collaboration is beneficial for the stock of scientific and technological knowledge. The evidence in favour of positive gains from collaboration and networking on scientific performance is mixed however.

Teams typically produce more frequently cited research than individuals do, with this difference increasing over time. Teams are now also intensive producers of exceptionally high-impact research, a domain that was once dominated by solo authors.

A snapshot of a network at a given moment in time is not useful as a means of gaining insights into the co-variation and arguably the direction of the causality between changes in the network and scientific performance.

2 Theoretical Framework

Across the literature, the word "network" has an element of ubiquity that depending on the discipline and focus of the study can have different meanings. As an object of analysis, attention has been paid to understanding the processes of network formation or to derive properties from network topologies. Other sets of inquiries have used the network as an analytical tool to understand social dynamics, to develop algorithms or to trace relationships between objects[1]. As an analytical framework, the network has served as a structure, channel and environment that theories have used to derive implications from social interactions. For instance, social capital refers to the availability of resources shared through the goodwill of individuals and reachable through a "structure" composed of social interactions (Adler & Kwon, 2002). The social cognitive theory proposes that behaviour and the cognitive biases of individuals can be derived from their social "environment", in other words, due to the interplay of individuals in a network (De Carolis & Saparito, 2006).

A different approach is to take the network as a theory to understand the relationship between a well-defined structure, given by a set of nodes with a specific type of connection, and certain outcomes (Borgatti & Halgin, 2011). For instance, the network framework, can explain how knowledge and resources are exchanged between agents (Haythornthwaite, 1996); also what are the patterns of scientific collaboration (Newman, 2001) and how the overall network structure affects process of knowledge dissemination (Cowan & Jonard, 2004). Different from social capital theory, network theory seems not impose ex-ante assumptions on why individuals exchange resources, such as assumptions based on "goodwill". Moreover, emphasis is not given to one specific consequence of being

[1] For instance, structural function of the brain(Bullmore & Sporns, 2009), or structural relationships between proteins (Amitai et al., 2004).

in a network, such as how the network affects cognition. Instead, it is the nature of the connection that defines the network and the kind of resources flow through social ties. Consequently, in order to draw conclusions from network theory it is necessary to define who the agents are and what is their connection type, as well as the kind of resources that are potentially flowing through the network in order to to infer conclusions from the network structure.

2.1 Research collaboration: flow of ideas, knowledge and resources.

There is no common definition of Research Collaboration (RC), mainly because as an object of analysis has many dimensions or facets that have been studied from different contexts (Bukvova, 2010). Collaboration, however, is perhaps rooted in the evolutionary trajectory of our species, is referred to as the human "ability to obtain otherwise inaccessible goals", and relies on mechanisms of coordination and resource distribution between partners (Melis, 2013). From the perspective of RC, there are three key elements from the previous definition of collaboration that are important to highlight, namely that it: 1) is related to a certain activity; 2) is based on partnership and 3) has a common purpose.

RC is a partnership defined by associations between scientists, researchers and in general groups of scholars working at the intersection or frontier of some discipline. The body of practitioners have a recognized expertise and competence within a domain or field which provide them with certain power and capacity to diffuse new ideas and information through networks referred as "epistemic communities" (Haas, 1992). The dissemination of information can induce new patterns of behavior and coordination between members of the community and RC is acknowledged as one of the main communication channels used by scholars to circulate knowledge. The flow of information is tied to processes of knowledge production where the community structure assigns rewards and gives recognition to members who are pioneers in their respective scientific fields (Stephan, 1996). Novelty is rewarded because is through processes of knowledge production that disciplines transform and RC as partnership based on knowledge exchange can play a relevant role in their evolution.

The process of knowledge production, is different from other forms of transformation because the means are composed by diverse blend of tangible and tacit resources. Many authors have pointed out a connection between process of knowledge production and diversity both in terms of ideas and individuals. The mere creative process behind knowledge generation relies on the capacity of making connections between seemly unrelated ideas (Simonton, 2004). Innovations and scientific discoveries have shown to be the product of accretion processes pushed more by collective effort rather than work of single individuals (Berkun, 2010) and highly cited publications are usually produced by a diverse group of researchers (R. B. Freeman & Huang, 2014). In a similar way the partnerships of RC has been described as the union of individuals that differ notably and share information to achieve a common goal (Amabile et al., 2001). RC is often associated with concepts of integration and cross-functional linkages that basically refer to a diversity of individuals coming together to achieve a mutual objective (Jassawalla & Sashittal, 1998). More specifically than diversity in terms of individuals, RC is also arranged around researchers exchanging a variety of competencies and resources (Melin & Persson, 1996). RC also plays a significant role in the dissemination of knowledge spillovers from R&D (Crespi & Geuna, 2008), not only at the national level but also through the sharing of the pool of multinational research knowledge among researchers working together. Resources for knowledge production are broader than simple funding for projects, and in the case of basic and applied science or larger scientific endeavours can include the sharing and pooling of instrumentation, machinery, access to laboratories and other materials.

RC is also a particular kind of social tie, drawn and enriched by the endowments of scientific and technical human capital, such as skills and knowledge (Bozeman & Corley, 2004) that encompass a broad spectrum of scientific exchanges and interactions, not all necessarily oriented or captured by the byline of an academic publication (J. Katz & Martin, 1997). Nevertheless, co-authorship has served as a well documented proxy for RC in a growing body of studies (Ajiferuke, Burell & Tague, 1988; Glänzel & Schubert, 2005) and knowledge production is an integral part of the work done by epistemic communities that leave a codified trace in academic publications. In this context RC is defined as the scientific or academic activity that encompasses the ability to create and maintain social ties to exchange ideas, knowledge, resources and skills for the purpose of communicating and generating new knowledge. The close link between RC and knowledge generation support the expectation of positive gains from the exchange of ideas and resources through scientific and academic social ties.

2.2 Network Structure and Scientific Performance

The core statement of a network approach to research collaboration is that knowledge production is not solely a function of individual effort, skills and endowments of knowledge but a share of inputs are exchanged through interaction. From the collaboration framework, researchers are conceived as interdependent stakeholders who hold control over resources, such as knowledge and expertise, and their association is based mainly on common interest when the potential benefits of working together are higher than individual returns(Kramer, 1990; Wood & Gray, 1991). RC leave a trace in academic articles when knowledge production occurs and these associations are represented as dyadic relations between authors whose paths compose the network structure. This structure is not static because authors' social connections can grow or decay over time, thus shaping the evolving space of interaction.

Analysis on the network structure is relevant because reflects the propensity of RC that affects knowledge production via *exchange* mechanisms, where researchers who perceive a benefit from their association, exchange a set of resources both tangible and from the diversity of research ideas. More importantly, from the pool of resources and human capital endowments from RC, authors also benefit from *synergy* mechanisms, where collective inputs can produce a total effect that is greater than the sum of individual contributions. For instance, epistemic communities thought collaboration can develop a common language and a set of rules that enhance the coordination and exchange by reducing the transactional cost. Thus, network positions provide an insight into how researchers interact and how resources and information flow through social ties.

The space of interaction is defined from a set of vertices or authors $V(G)$ and a set of edges or connections $E(G)$ that constitute a network $G = (V, E)$ and take the form of square n-dimensional adjacency matrix, $A(G)$ whose off-diagonal elements constitute connections between authors, v_i, $v_j, \ldots, v_n$, such that $a_{ij} = 1$ if $(v_i, v_j) \in E$ and zero otherwise. From this structure, the focus of the study lies on assessing the relationship between individual level network positions and scientific performance. Firstly, there are measurements developed to a the analyze level of influence, access to resources, brokerage and information spreading, that mostly come from a long tradition of works on network centrality (Sabidussi, 1966; Bonacich, 1972; L. Freeman, 1978; Borgatti, 2005). Secondly, there is a group of network metrics from the literature of social capital that pays attention to the network structure in relation to closure in social groups and redundancy of information from social ties (Burt, 2001).

- *Degree*: Is a measure of the propensity of collaboration, reflects the direct connections from an author i to his co-authors j. Researchers with a higher degree in comparison to other authors can benefit from holding a strategic position which allows them to exchange resources and influence their epistemic communities.

$$D_i = \sum_{i \neq j \in V} E(v_i, v_j)$$

- *Closeness*: Accounts for the social distance between an author and all the other researchers in the network. Social distances or paths are measured by the adding of all edges between two vertices and the shortest path separating to vertices is called geodesic distance $d(v_i, v_j)$. This measurement does not capture mechanisms of exchange from direct connections, but a close distance to all other authors reflects privileged access to information and valuable knowledge regarding the expertise and endowments of other potential co-authors.

$$C_i = \frac{1}{\sum\limits_{i \neq j \in V} d(v_i, v_j)}$$

- *Betweeness*: Reflects the author's brokerage capacity by measuring the intensity from which his network position lies in the geodesic path of other researchers in the network. Researchers working as bridges between communities can gain from a diverse set knowledge and resources from connections to different epistemic boundaries.

$$B_i = \sum_{i \neq j \neq k \in V} \frac{g(v_j, v_k)|v_i}{g(v_j, v_k)}$$

- *Eigenvector*: This network measurement takes into account the heterogeneous influence from connections to well-connected co-authors. Researchers with a high eigenvector rank can to gain from mechanisms of exchange from connections to resourceful co-authors that are careers of knowledge and resources from their communities. If e_i is the rank for each researcher, where d_j is the degree of each neighbour and λ is a constant:

$$E_i = \lambda^{-1} \sum_j A_{ij} d_j$$

By defining the vector $\mathbf{e}$ whose elements are e_i, then the derivation of the rank is homologous to finding a solution to the eigenvector problem for the adjacency matrix $A(G)$.[2]

$$\mathbf{Ae} = \lambda \mathbf{e}$$

- *Reach*: Considering that information and resources travel efficiently in the shortest path separating two authors in the network, the longest geodesic distance between two authors describes capacity to disseminate and receive information from his community. Reach is defined as the shortest geodesic path between an author i and all the other researchers of his community.

$$R_i = \max_{i \neq j \in V} d(v_i, v_j)$$

- *Transitivity*: Is a measurement that captures the effect of researcher's membership to an interactive epistemic community. An interconnected network neighbourhood is related to closure in research groups where clusters of collaboration can facilitate the coordination and exchange between authors. Transitivity for an author is operationalized as the proportion pairwise co-authors that are also connected among themselves. A clique is a part of network (sub-graph) where each pair of vertices is adjacent and transitivity is defined as the ratio of the number of size 3 cliques and the number of vertices connected by two two edges, or triplets, that fall on an author v_i.

$$T_i = \sum_{i \neq j \neq k \in V} \frac{c(v_j, v_k)|v_i}{t(v_j, v_k)|v_i}$$

- *Structural Holes*: A key idea behind the concept of structural holes is that a source of novel research ideas come from connections to a diverse set of epistemic communities working relatively separated from one another. Researchers working in close proximity, can reduce transactional cost and improve communication between them but as the level of interaction increases such as the overlap of research ideas. Thus, connections to otherwise disconnected communities provide opportunities for accessing to seemingly unrelated ideas that can serve authors to develop new insights into his research. Since collaboration requires a certain investment of time and effort to sustain a mutually beneficial exchange, links to researchers with mutual connections increase the overlap of information and constrain the potential access to structural holes. The index of constraints to structural holes for an author is measured from an index that accounts for the proportion of direct and indirect connections to other authors.

$$H_i = \sum_j C_{ij}$$

Each element of the constraint matrix C_{ij} is given by the sum square of the proportion of direct connections $p_{ij} = z_{ij} / \sum_q z_{iq}$ and the product of the proportion of mutual contacts $p_{iq} p_{jq}$.[3]

$$C_{ij} = (p_{ij} + \sum_q p_{iq} p_{jq})^2$$

2.3 Relational attributes of the network in research collaboration

Traditionally, network theory as a "model of flow" relies on the structure alone to measure social distances from paths between authors to derive individual-level network structural characteristics,

[2] Undirected networks take the form of a square matrix $A(G)$ with non-negative values $a_{ij} >= 0$ which yield a unique non-degenerate eigenvector for largest eigenvalue that can be computed by an iterative algorithm (Spizzirri, 2011).

[3] The strength of the connection between two vertices z_{ij} is given by some sort of weight, for instance, an intensity of collaboration between two authors. However, for an unweighted network, the strength is the same for all authors, and the source of variation is the degree and overlapped co-authors.

discussed in section (2.2), and estimate their effect on outcome variables. However, analysis of network structure usually does not take into account individual attributes or other contextual factors that can be relevant to understand the behaviour of outcome variables (Borgatti & Halgin, 2011). For instance, in the context of RC, analysis based solely on the network structure can not take into account researchers' attributes for example age, academic rank, and other unobservable characteristics such as ability and skills that might affect scientific performance. But, more importantly analysis based solely on network structure can overlook a relational dimension of RC that is relevant to disentangle the quality and quantity of flows circulating through collaboration ties. These relational attributes are for instance, the *institutional*, *disciplinary* or *geographical* dimensions of RC (Amabile et al., 2001; Sonnenwald, 2007; Bukvova, 2010). The relational dimension of networking complements the network structural positions because add more layers of information about the quantity and quality of network flows. Therefore, in addition to the effect of the network structure this study pays attention to the effect of disciplinary knowledge flows; positive externalities related to flows of knowledge and resources from connections to academic institutions; and flows from connections to international hubs of knowledge production.

- *Institutional Collaboration:* From an index set of individuals $\{1, 2, \ldots, n\}$ is defined a set of independently distributed random variables $A_{[n]} = \{A_1, A_2, \ldots, A_n\}$ of connections to academic institutions. Each random variable $A_{[n]}$ takes values from a finite set of academic institutions X_i given by the institutional affiliations of an individual i in addition to the affiliations from his co-authors. Is expected that each academic institution of X_i embody a set of positive externalities from collaboration in terms of knowledge and access to resources, thus authors can gain from the diversity of connections to academic institutions. An index of *multinstitutional collaboration* for each author i is defined by the measuring the diversity of connections to academic institutions given by the cardinality of the pair-wise disjoint set from the realizations of each random variable in $A_{[n]}$, such that for each i author:

$$MI_i = |X_j \cap X_k| : (X_j \neq X_k) \in A_i \tag{1}$$

  The multinstitutional index captures the range or the breadth by which researchers can disseminate and access knowledge and resources from a diversity of academic institutions. The use of the cardinality of a disjoint subset of the realizations from connections to academic institutions has a meaning in the light that more diversity in the realization set of $A_{[n]}$, represents less redundancy of knowledge and resources from each X_i in the subset. However, from the diversity alone, is not possible to evaluate how the different academic institutions are interrelated, in other words, what is the frequency from which an author had a link to an academic institution and not the other, and how the allocation of connections to academic institutions can affect scientific performance.[4] Authors with less frequent and more range of connections to different academic institutions can benefit from having a more interconnected set of $A_{[n]}$. A highly interconnected set of connections to academic institutions can confer advantages to authors because they can combine the knowledge and ideas from different sources, connect groups across universities and other research facilities, but also because they can potentially benefit from funding of projects and other institutional level grants. An author level index of *interinstitutional collaboration* measures the interconnections between academic institutional connections by applying Shanon Entropy on the set of realizations $A_{[n]}$. The level of entropy[5] reflects the degree of institutional interconnection because the function grows when the distribution of set of realizations $A_{[n]}$ has less frequent connections or becomes more uniformly distributed while the diversity of academic connections, MI_i, increases. The index is defined as follows:

$$II_i = -\sum_{x \in X_i} p(A_i = x) \log_2 p(x) \tag{2}$$

- *Disciplinary Collaboration:* A discipline represents a body of knowledge centred around certain topics or areas of study with a degree of distinguishable recognition from other fields which make possible to systematize or categorize the structure of science. A discipline has also

[4] For instance, researcher 1 has a realization set of $A_1 = \{a, a, a, a, a, a, b, c, d\}$ and $MI_1 = |\{a, b, c, d\}| = 4$, but researcher 2 with a more uniform distribution of $A_2 = \{a, a, b, b, c, c, d, d\}$, will also have a $MI_2 = 4$

[5] The measure of entropy has been used primarily to measure the average uncertainty or informational content, by using a binary logarithm to represent the number of bits or the number of steps in a decision tree required to describe a random process.

differentiated practices, methods, norms and forms of communication with defined terms or jargon that shape the behaviour of practitioners approaching an object of study, in many ways disciplines are or can be taken as institutions (Buanes & Jentoft, 2009). But disciplines as bodies of knowledge or as social structures shaping and constraining behaviour do not exist independently of the practitioners that conceived them, legitimize and expand their frontiers, in other words, disciplines are tied to their epistemic communities. These communities, produce knowledge in a more specialized way than in the past, articulated in more disciplines, where both academic and social challenges are becoming more complex, pushing changes in the organization and communication of science in such a way that the disciplinary boundaries are becoming less clear (Gibbons et al., 1994). The disciplinary interconnection is rooted in the combination of ideas to induce novelty and creativity, where researchers are able to break through limiting beliefs, bring fresh perspectives, fit disciplinary holes, reduce the risk of narrow vision and their implicit costs (Nissani, 1997). Changes in disciplinary collaboration patterns, can come from the evolution of institutional arrangements of disciplines, for instance, the rise of multidisciplinary projects, openness of science and disciplinary migration, or changes related to self-organization of authors, that are out of the scope of this study to observe. Nevertheless, from the network perspective, the disciplinary interconnection will be given mainly by the exchange of knowledge and ideas from different bodies of knowledge by means of RC. Thus, in order assess knowledge flows from interaction a useful framework is to characterize the disciplinary interrelation of RC in terms of knowledge diversity and coherence (Rafols & Meyer, 2010; Liu, Rafols & Rousseau, 2012). "Knowledge diversity" refers to the breath or range of use from a diverse set of bodies of knowledge, and in the context of knowledge flows from networking reflects the content of disciplinary knowledge potentially accessible thought RC. Derived from the definition of knowledge diversity, *multidisciplinary collaboration* is defined for this study as the extent from which researchers can use directly or through peers different bodies of knowledge to find novel and creative approaches to tackle common issues. Thus, the degree of multidisciplinarity in this study is not only given by the disciplinary specialization of individual researchers but also from the diversity of disciplinary specializations accessible through RC. From the previous definition a set of independently distributed random variables $D_{[n]} = \{D_1, D_2, \ldots, D_n\}$ of disciplinary connections is defined, all taking values from a set of disciplines W_i.[6] Then, similarly to equation (1), a multidisciplinary index of connections is given by the cardinality of the pair-wise disjoint set from the realizations of each random variable in $D_{[n]}$, such that for each i author:

$$MD_i = |W_j \cap W_k| : (W_j \neq W_k) \in D_i \tag{3}$$

"Knowledge coherence" is related to how the bodies of knowledge are used and how they are interrelated. Different from knowledge diversity of multinstitutional collaboration which reflects the divergence of disciplinary knowledge flows, knowledge coherence reflects their convergence. For instance, from each disciplinary realizations set $D_{[n]}$, divergence is evaluated by the unique set of bodies of knowledge accessible for each researcher through RC in MD_i. Then similar to equation (2) which reflects the interrelation between connections to academic institutions, if the degree of convergence between disciplinary knowledge flows is high then the distribution of a realization set in $D_{[n]}$ will be more homogeneous or uniformly distributed[7], and as the divergence of different bodies of knowledge increases (MD_i) such as the level of entropy. *Interdisciplinary collaboration* is defined as the degree from which researchers can integrate a diverse set of disciplinary bodies of knowledge accessible through RC, and the index is given by:

$$ID_i = -\sum_{w \in W_i} p(D_i = w) \log_2 p(w) \tag{4}$$

- *International Collaboration:* The growth and the effect of international collaboration on scientific performance are in a way counter-intuitive because the exchange of codified knowledge in the form of data, ideas or publications and face to face interactions have lower transaction-cost in comparison to long-distance exchanges.[8] But contrarily to the expectation, country-level

[6] The disciplines are defined using 252 set of Web of Science Categories field.

[7] Notice that if a disciplinary set of knowledge flows in $D_{[n]}$ has full converge, then all the knowledge production of a researcher and his collaborators will be specialized in one discipline, then $MD_i = 1$ and $ID_i = 0$.

[8] For instance, empirical studies have shown that the probability of collaboration (tie formation) decreases with the geographic distance separating authors.

analysis of knowledge production have shown that during the past three decades international collaboration has been growing and forming dense networks with defined knowledge hubs (Leydesdorff & Wagner, 2008) and that internationally authored publications can largely be driving the growth of knowledge production of nations (Adams, 2013). Additionally, it appears that the growth of international collaboration is not independent of disciplines that rely more on codified knowledge (Logic and Mathematics) than on resources (Astrophysics and Polymers), which poses the possibility that the gains could be related to informational gains from diversity of ideas from researchers working across geographic-intellectual boundaries (Wagner, 2005). Similarly to access to structural holes from the network structure, connections to international hubs of knowledge are carriers of novel information from researches working at the frontier of disciplines with a degree of geographical separation. An index of international collaboration intensity is defined as the sum of all international connections from a set of vertices v_n of national authors $n \in N$, and a set of vertices of international co-authors $c \in C$.

$$IC_i = \sum_{i \in N, c \in C} E(v_n, v_c) \tag{5}$$

Authors working at the intersection between national and international networks have a special network characteristic referred as "gatekeepers" because they can connect the local and foreign groups (Tichy, Tushman & Fombrun, 1979) of researchers and function as brokers of knowledge and resources between them (Gould & Fernandez, 1989). For instance, gatekeepers can benefit from accessing a pool of multinational spillovers from R&D and hold a power position because they can constrain or mediate the connections between the local and external network. A *gatekeeping index* is defined as the ratio of international and local connections:

$$GI_i = \frac{\sum\limits_{c \in C} E(v_i, v_c)}{\sum\limits_{n \in N} E(v_i, v_n)} \tag{6}$$

3 Methodology

Previous studies have identified many challenges associated with author level estimations of the effects of collaboration on scientific performance, examples being name disambiguation, cohort effects and the assortativity of social networks (Kumar, 2015). Name disambiguation is a common challenge for studies that rely on the most frequently used bibliometric databases. The capacity to consistently track the career of authors is a major concern that can lead to questions about the reliability of some estimations. To address the disambiguation issue a special dataset of Mexican authors provided information on the publication curriculum of authors. The use of publication records granted an advantage over other disambiguation methods that rely on the co-citation network, which is weak at tracking authors across disciplines and time (Schulz, Mazloumian, Petersen, Penner & Helbing, 2014), as well as relying on the authors' names on the publication byline. The process of matching the publication records with Web of Science was based on an algorithm that used regular expressions on the title, journal name, doi and other paper level meta-data. In addition to the matching process, an institutional name disambiguation algorithm was implement based on a set of rules that used string analysis for measuring word similarity and editing distance, that has been shown to be precise (S. Huang, Yang, Yan & Rousseau, 2014).

Another concern for the quality of the estimation are cohort effects on scientific performance, in particular, age, discipline and academic rank of authors. Previous studies have shown a convex pattern between publication output and age (Rørstad & Aksnes, 2015), also confirmed in a study that used a related dataset of Mexican authors (Gonzalez-Brambila & Veloso, 2007). The implication of this pattern is that researchers reach a peak in their productivity at some point in their careers and after they tend to produce less. Age can also be used as a proxy for academic rank, mainly because the experience is accumulated through time, but also because there is a low expectation of young researchers with full academic professorship. A second issue is the large differences in scientific performance not inherited from collaboration but from the differences in publication and citation rates across disciplinary fields (D. Hicks, 2005), that invalidates the use of central tendency measurements, absolute values and other bibliometric indicators to perform a fair cross-disciplinary comparison of productivity or citation impact.[9] A suggested solution is to apply normalized scientific performance measurements, defined in section (4.1), that follow the intuition of comparing

[9]For instance, the field of Biology reported a yearly average of more than 7000 citations, whereas the field of

individual-level publication and citation outputs to the output of other researchers in their respective fields. Failing to normalize the dependent variables can introduce large bias on the network variables estimates, i.e, potentially attributing a network effect what is corresponding merely to the distribution of publications and citations of the publishing field of authors, similar to comparing apples and oranges (Abramo, Cicero & D'Angelo, 2013). Impact and publication rates can also be a function of academic rank and carrier cycles that can be difficult to control. It is likely that senior researchers will be more productive and visible when compared with younger scholars, though scholars that take on administrative positions or jobs outside academia may have a less frequent publication rate. To account for these issues a career control was implemented to measure cumulative publication effort (Ductor, Fafchamps, Goyal & van der Leij, 2014).

The main estimation issue in this study is the potential bias of estimates due to endogeneity of the network variables. The econometric concern arises because individuals make deliberate choices to collaborate based on time-varying unobserved characteristics that may be correlated with scientific performance. Attributes related to idiosyncratic attributes that influence the likelihood of engaging in collaboration, for example, homophily between researchers, working style and degree of extroversion, are assumed to be time-unvarying. However, there is also some potential time-varying heterogeneity, in the form of external incentives, such as funding, reputation, institutional agendas and projects, that affect the likelihood of collaboration because they require some complementarity of knowledge and skills (Hara, Solomon, Kim & Sonnenwald, 2003). These external incentives can be conceived as "opportunity sets" in which authors apply or receive project offers from which simple ideas are sole-authored, whereas more complex projects that require complementarity in knowledge and skills can be co-authored.

Aside from a experimental design, there are essentially two approaches that have been applied to address the endogeneity of network variables (Blume, Brock, Durlauf & Ioannides, 2011; Horrace, Liu & Patacchini, 2016). The first approach assumes that network formation occurs in a two-step process, with individuals in the first stage self-selecting in to a group based on network-specific characteristics, and in a second stage individuals choosing to collaborate or create a link based on individual characteristics that are potentially unobserved. This approach assumes that once the potentially unobserved time-varying factors are driving the formation of links are addressed[10], the network variables can be treated as exogenous. The second approach assumes that network formation is a function of individuals making decisions, and use structural modelling to account for the individual selection bias (Goldsmith-Pinkham & Imbens, 2013). The advantage of the latter approach is that the estimates are produced by explicitly modelling the process of network formation with a link into the output equation. However, these methods are more developed for a static setting, where the network is represented as a snapshot of a span of years, and has not time dimension (Hsieh & Lee, 2017), and cannot be easily applied to a dynamic network process in a panel setting. Therefore, the applied estimation strategy is in line with the former approach and based upon previous methodologies (Bramoullé, Djebbari & Fortin, 2009; Ductor, 2015). In particular, I combine information on the structure of the network from the past and external exogenous variables to derive a set of instrumental variables based on the commonality of research interests.

4 Econometric Framework

In line with the previous studies (Claudia N. Gonzalez-Brambila, Veloso & Krackhardt, 2013; Ductor, 2015; Rodriguez Miramontes & Gonzalez-Brambila, 2016), this analysis assumes that the impact of network effects of collaboration on scientific performance is a function of their recent past, in other words, the exchange of ideas, knowledge and resources between authors due to collaboration will pay-off in the immediate future. This is a valid assumption considering that collaboration is not necessarily co-authorship, therefore, the exchanges and the social link are not only bounded to the time period in which researchers share a byline, but the effect of interaction is present on the immediate future. Additionally, a non-contemporaneous estimation in a panel setting can break the latent effect of feedback loops between scientific performance and collaboration. Further, I expect that the knowledge production of authors is better captured by measuring the total publication output at intervals of time. This is because not all authors publish on a yearly basis, but also because the span of projects and editorial times usually can take longer than one year. Therefore, 25 years of publication records of authors from 1998 to 2013 are used to capture the co-variation

Industrial Engineering received around 83 yearly average citations. Similarly, the field of Astronomy and Astrophysics had a yearly average publication output of 265, and the Anthropology field published on average 13 papers per year.

[10]Studies that have used instrumental variable approach, GMM or Heckman Models.

between network variables and scientific performance. The network variables were estimated at t intervals of $[y-1, y]$, giving a total of 11 periods. The dependent variables were constructed in the immediate future, such that publication productivity in each period was measured in intervals of two years $P_{it} = \sum_{n=1}^{2} f(y_n)$. In a similar fashion the citation impact was computed in fixed intervals of 3 years after publication $C_{it} = \sum_{n=1}^{2} f(y_n, y_{n+2})$.[11]

4.1 Measurements of Normalized Scientific Performance

The normalized measurements of scientific performance for publication productivity and citation impact used to estimate the dependent variables are defined follow:

- **Normalized Citation Impact (NCI)**: The citation score is defined for a set of i authors, a set of k publications, a set of j subject areas of Web of Science and a set c of citations. For k publications in the sample, the mean ratio of the expected citations in j disciplines is calculated.

$$NCI_k = \frac{\sum_{j=1}^{n} \frac{c_k}{E[c_j]}}{n}$$

The Normalized Citation Impact is defined as the average of NCI_k for each k publications authored or co-authored by an author i.

$$NCI_i = \frac{\sum_{k=1}^{n} NCI_k}{n}$$

- **Discipline Weighted Publication Productivity (DWPP)**: Is an index developed to normalize the publication output of authors accordingly to the publication intensity in each discipline (Yamamoto & Ishikawa, 2017). For each author i a set of effort ratios per discipline is defined, such that k_{ij} are the number of publications of author i that belong to a j subject areas of Web of Science and k_i is the total number of publications of author i.

$$E_{ij} = \frac{k_{ij}}{k_i}$$

Further, disciplinary productivity is defined as a ratio of the total publication output and total effort for each j subject area of Web of Science, integrated over i authors.

$$P_j = \sum^{i} \frac{k_{ij}}{E_{ij}}$$

For an author i the DWPP is measured as the weighted sum of total number of publications in each j discipline weighted by the inverse of the disciplinary publication productivity.

$$DWPP_i = \sum^{j} \frac{k_{ij}}{P_j}$$

4.2 Econometric Estimation

The main estimation concern is the potential endogeneity of network variables. In order to obtain unbiased estimators, an instrumental variable strategy was applied based on the derivation of commonality of research interests using information of past network structure. Is expected that a key driver affecting the likelihood of collaboration is the similarity of authors research interest (Fafchamps, van der Leij & Goyal, 2010). Authors can have a high degree of homophily or other unobservable attributes that affect the likelihood of collaboration, however, without topics in common is unlikely that will work together. A commonality of research interest is a valid instrument because can potentially drive the formation of the network and will only affect measurements of scientific performance through network variables.

[11] In order to measure the effect of the network and not the gains of impact due to the passage of time, a fixed citation window was used for all publications.

A proxy of the overlap of research interest between an author and his potential co-authors can be derived using information from the structure of the network and vectors of disciplinary specialisation (Fafchamps et al., 2010; Ductor, 2015). From an author's publication output, a vector of disciplinary specialisation, S_{ij} , is composed by representing each Web of Science Category in a j-dimensional space. The direction of this vector represents the research interest of an author in a period. Similarly, using the network structure, a publication output of co-authors' coauthors, D^2_{ij}, is used to estimate a vector of disciplinary specialization of an author's potential co-authors. Given that these two vectors are relative good approximations of the research interest of an author and his potential co-authors, a cosine function is used to measure the commonality of research ideas. If an author's vector of disciplinary specialization is approximately orthogonal to the vector of disciplinary specialization of potential co-authors the cosine similarity is close to zero. Conversely, if the two vectors of disciplinary specialization have the same direction, the commonality of research ideas will be close to one. As mentioned in studies of Ductor and Fafchamps et al., is possible that the likelihood of collaboration and the commonality of research ideas follow a concave shape function, therefore the quadratic form of commonality of research ideas is also exploited as an instrument. The reason behind is that a high degree of similarity of research ideas can discourage collaboration because there is no complementary of skills and knowledge. The commonality of research ideas was estimated in each t from past network structure in the following way.

$$C_i = \frac{\sum^{j} S_{ij} D^2_{ij}}{\sqrt{(\sum^{j} S_{ij})^2 (\sum^{j} D^2_{ij})^2}}$$

For simplicity of notation variables of network structure and collaboration variables, described in Section (2.3) and (2.2), can take the form of column vectors n_{itl} and c_{itm} respectively. The interest lies in the estimation of coefficients β for network variables and γ for collaboration variables. A $log(y + c)$ transformation was applied on both dependent variables for the purpose of treating non-linear behaviour between the dependent variables and regressors and also as a mean to account for observations with zero values on the measurements of scientific performance. The main estimation equation is the defined as:

$$log(y_{it} + c) = \alpha + n'_{it}\beta + c'_{it}\gamma + v_{it} \tag{7}$$

The model has a one way error model, $v_{it} = \mu_i + u_{it}$, where μ_i is a vector of individual time-unvarying unobserved characteristics and u_{it} is a vector of non-serially correlated idiosyncratic errors. Furthermore, because is likely that u_{it} will be correlated within groups of i researchers across time, robust clustered errors were used. Network variables are represented as a n vector $1 \times l$, and collaboration variables are a vector c of dimension $1 \times m$. The total number of regressors is given by $k = l + m$. Because we have l potentially network endogenous variables and only two instruments, is not possible to produce a joint estimation of network and collaboration variables using an instrumental variable two-stage least square (2SLS-IV) procedure. Nevertheless, for robustness this estimators are presented for each of the network variables in table **X**. The proposed solution is to estimate the model using a two-stage system Generalized Method of Moments (GMM), that is robust against heteroskedasticity and potential endogeneity of regressors. GMM was original conceived to exploit the set of lagged differences of endogenous and other exogenous variables as instruments in a panel setting (Arellano & Bond, 1991; Blundell & Bond, 1998). The selected system GMM, in addition to the lagged differences added lagged levels, as internal instrument or "moment conditions", that potentially increase the performance of the estimation (Blundell, Bond & Windmeijer, 2001). The main assumption of the model is that lagged differences and levels are valid instruments since they are good predictors of themselves and uncorrelated with the u_{it}. Additionally to the moment conditions used by the system GMM, two "external instruments" were included in the estimation, the commonality of research interest and also its quadratic form. Because is likely that both internal and external instruments will be correlated with the time-unvarying observables μ_i, a useful transformation is to take the first difference the system, therefore equation (7) takes the form of equation (8) which is used to estimate coefficients on both citation score, NCI, and publication productivity $DWPP$.

$$log(y_{it} + c) - log(y_{it-1} + c) = \alpha + (n_{it} - n_{it-1})'\beta + (c_{it} - c_{it-1})'\gamma + (\mu_i - \mu_i) + (u_{it} - u_{it-1})$$

Rewriting

$$\Delta log(y_{it} + c) = \alpha + \Delta n'_{it}\beta + \Delta c'_{it}\gamma + \Delta u_{it} \tag{8}$$

5 Descriptive Statistics

From the network perspective, each collaboration has at least one corresponding edge and two vertices, that represent authors. Essentially, structural variables of the network are obtained by counting in different ways the social distances between authors derived from the $A(G)$ adjacency matrix. In some cases, for instance, if the network structure is close to a fully connected graph, network measurements will yield identical values. Previous studies have found a high correlation between the network structure measurements (Marsden, 2002; Valente, Coronges, Lakon & Costenbader, 2008; Yan & Ding, 2009). This can be problematic because the amount of novel information that network variables can add in an econometric setting will be questionable. This was not the case for this analysis, as reported in Table (1) the correlation between network variables for our sample is low; in addition, the correlation of collaboration variables is provided in Table (2).

	1	2	3	4	5	6	7	8	9	10
1. YNCI										
2. YDWPP	0.21***									
3. PUBS	0.11***	0.63***								
4. CITES	0.18***	0.24***	0.36***							
5. CLOSENESS	0.21***	0.02***	-0.01**	-0.03***						
6. DEGREE	0.17***	0.24***	0.28***	0.42***	-0.03***					
7. BETWEENNESS	0.02***	0.32***	0.48***	0.17***	-0.09***	0.19***				
8. EIGEN_CENTRALITY	0.02***	0.05***	0.07***	0.03***	0.09***	0.04***	0.00			
9. STRUC_HOLE	-0.05***	-0.25***	-0.38***	-0.14***	0.07***	-0.19***	-0.32***	0.00		
10. TRANSITIVITY	-0.04***	-0.26***	-0.42***	-0.07***	0.01	-0.03***	-0.25***	-0.01	0.30***	
11. REACH	0.02***	0.08***	0.15***	0.02***	0.03***	0.03***	0.03***	0.00	-0.22***	0.08***

*p<.05, **p<.01, ***p<.001

Table 1: Scientific Performance vs Network Variables

	1	2	3	4	5	6	7	8	9
1. YNCI									
2. YDWPP	0.21***								
3. PUBS	0.11***	0.63***							
4. CITES	0.18***	0.24***	0.36***						
5. N_INST	0.17***	0.36***	0.49***	0.53***					
6. E_INST	0.07***	0.27***	0.40***	0.24***	0.67***				
7. N_DISC	0.06***	0.40***	0.65***	0.15***	0.32***	0.34***			
8. E_DISC	0.05***	0.30***	0.50***	0.11***	0.25***	0.31***	0.92***		
9. INT_RATIO	0.09***	0.05***	0.06***	0.13***	0.19***	0.24***	-0.02***	-0.03***	
10. N_INT	0.09***	0.20***	0.24***	0.55***	0.53***	0.15***	0.04***	0.02***	0.08***

*p<.05, **p<.01, ***p<.001

Table 2: Scientific Performance vs Collaboration Variables

By splitting the sample by quantiles, for both NCI and $DWPP$, Figures (1) and (2), show the different network patterns for researchers across different levels of scientific performance. Is clear that more productive and highly cited authors have as well a higher measurement of degree and betweenness centrality, however lower clustering coefficient than researchers from the first quantile. Is possible that more productive and highly cited researchers can make better use of resources from the network because they have more capacity to manage their social capital. For instance, they could have developed skills to coordinate projects or they could start connections faster because of establishment in their respective fields. This implies that the cost of creating and maintaining links can be lower for these group and in turn, can confer them advantage of collaboration strategies, i.e., complementary of knowledge and skills or division of labour. To test this hypothesis, the model from equation (8), is estimated for both dependent variables in each corresponding quantile.

6 Results

First, a set of regressions are estimated from Equation (6), without instrumenting the network variables. F

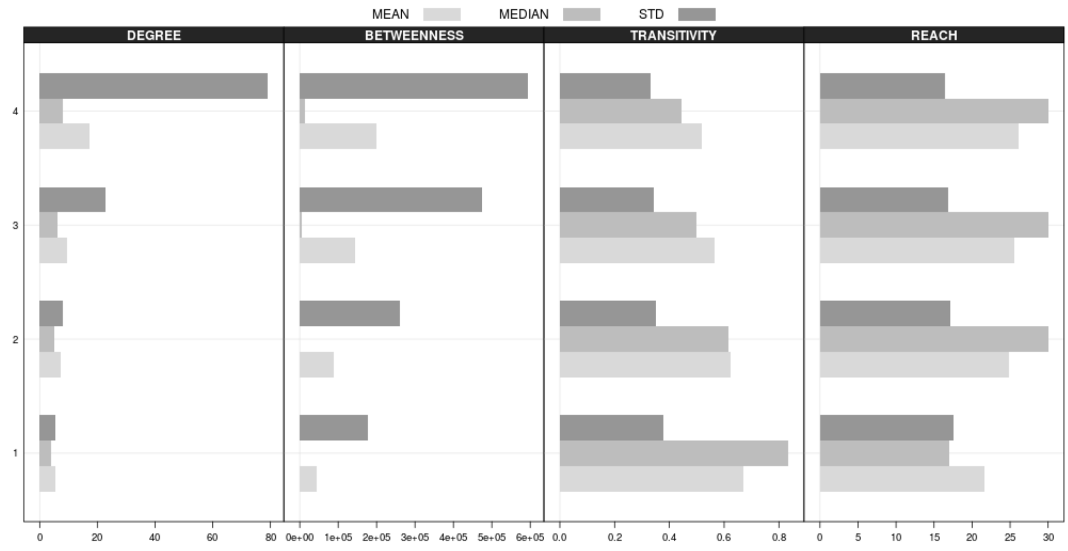

Figure 1: Quantile Statistics of Network Variables

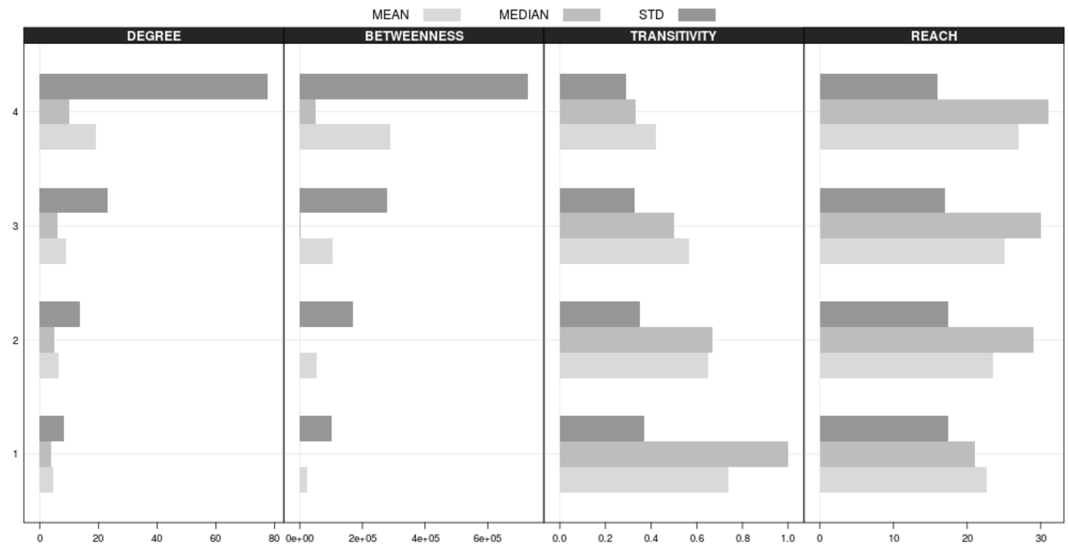

Figure 2: Quantile Statistics of Collaboration Variables

	Dependent variable:		
	NCI		
	Model 1	Model 2	Model 3
Publications	0.107***	0.107***	-0.0346***
	(-0.00929)	(-0.00916)	(-0.00766)
Age	-0.0617***	-0.0617***	0.248***
	(-0.00542)	(-0.00526)	(-0.0743)
Closeness	0.0841***	0.0841***	0.0860***
	(-0.0244)	(-0.0231)	(-0.0247)
Degree	0.0481***	0.0481***	0.0534***
	(-0.0106)	(-0.0103)	(-0.0113)
Betweenness	-0.0126*	-0.0126*	0.0051
	(-0.00644)	(-0.00665)	(-0.00447)
Transitivity	$7.98E-05$	$7.98E-05$	0.0088
	(-0.00557)	(-0.00546)	(-0.00599)
Structural H.	-0.0492***	-0.0492***	0.0145**
	(-0.00609)	(-0.00599)	(-0.00651)
Eigenvector	-0.0137*	-0.0137**	-0.0163**
	(-0.00699)	(-0.00693)	(-0.00715)
Reach	-0.0412*	-0.0412*	-0.0759***
	(-0.0234)	(-0.0222)	(-0.0236)
Multinst	0.0789***	0.0789***	0.0385***
	(-0.0112)	(-0.0114)	(-0.0142)
Interinst	-0.0223***	-0.0223***	-0.00978
	(-0.00779)	(-0.00777)	(-0.00984)
Multidisc	-0.0278**	-0.0278**	0.00946
	(-0.0138)	(-0.0134)	(-0.0123)
Interdisc	0.0592***	0.0592***	-0.0114
	(-0.0118)	(-0.0115)	(-0.0118)
Gatekeeper	0.0806***	0.0806***	-0.00429
	(-0.0051)	(-0.00506)	(-0.00599)
International	-0.0301***	-0.0301***	0.0232
	(-0.0076)	(-0.00722)	(-0.0153)
Year F.E.	No	Yes	Yes
Individual F.E.	No	No	Yes
Observations	48335	48335	48335
R-squared	0.068	0.106	0.042
Number of id_cvu			13727

	Dependent variable:						
	Normalized Citation Index						
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Publications	-0.0360^{***}	-0.0596^{***}	-0.139^{***}	0.046	-0.0492^{***}	-0.00235	-0.0359^{***}
	(-0.00751)	(-0.0135)	(-0.0473)	(-0.0362)	(-0.00952)	(-0.0789)	(-0.00751)
Age	0.270^{***}	0.141	0.287^{***}	0.361^{***}	0.205^{***}	0.276^{***}	0.272^{***}
	(-0.0744)	(-0.0865)	(-0.0789)	(-0.0896)	(-0.0767)	(-0.0884)	(-0.0745)
Closeness	0.0378^{***}						
	(-0.0138)						
Degree		0.316^{***}					
		(-0.119)					
Betweenness			0.348^{**}				
			(-0.152)				
Transitivity				0.229^{**}			
				(-0.104)			
Structural H.					-0.102^{***}		
					(-0.0355)		
Eigenvector						0.336	
						(-0.921)	
Reach							0.0348^{***}
							(-0.0131)
Multinst	0.0765^{***}	-0.139^{*}	-0.00137	0.0696^{***}	0.0726^{***}	0.0525	0.0771^{***}
	(-0.0141)	(-0.0838)	(-0.0386)	(-0.0136)	(-0.0142)	(-0.0652)	(-0.0141)
Interinst	-0.0316^{***}	0.0560^{*}	-0.0167	-0.0210^{**}	-0.0483^{***}	-0.011	-0.0311^{***}
	(-0.0101)	(-0.033)	(-0.0125)	(-0.00983)	(-0.0129)	(-0.0383)	(-0.0101)
Multidisc	0.0162	0.0205	-0.112^{*}	-0.0084	0.0259^{**}	-0.00158	0.0168
	(-0.0123)	(-0.0132)	(-0.0571)	(-0.0158)	(-0.0131)	(-0.0408)	(-0.0123)
Interdisc	-0.0199^{*}	-0.0336^{**}	0.0686^{*}	0.045	-0.0418^{***}	-0.0108	-0.0200^{*}
	(-0.0118)	(-0.0138)	(-0.0395)	(-0.0298)	(-0.015)	(-0.0168)	(-0.0118)
Gatekeeper	0.00208	0.0015	0.00481	-0.000765	0.0129	0.0101	0.00159
	(-0.00632)	(-0.00643)	(-0.00754)	(-0.00627)	(-0.00848)	(-0.0387)	(-0.00628)
International	0.0229	-0.0266	0.0497^{*}	0.0067	0.0226	-0.143	0.0225
	(-0.0145)	(-0.0342)	(-0.0292)	(-0.0156)	(-0.0141)	(-0.449)	(-0.0145)
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	48335	48335	48335	48335	48335	48335	48335
R-squared							
Number of id_cvu	13727	13727	13727	13727	13727	13727	13727

	Dependent variable:				
	NCI				
	Model 1	Model 2	Model 3	Model 4	Model 5
Publications	-0.0346^{***}	-0.0643^{**}	-0.0452^{**}	-0.0229	-0.00889
	(-0.00766)	(-0.0282)	(-0.0216)	(-0.0164)	(-0.0096)
Age	0.248^{***}	0.401^{***}	0.259^{*}	0.152	0.228
	(-0.0743)	(-0.153)	(-0.142)	(-0.144)	(-0.144)
Closeness	0.0860^{***}	0.178^{***}	0.184^{***}	0.0626	0.0566
	(-0.0247)	(-0.0557)	(-0.0555)	(-0.0516)	(-0.0394)
Degree	0.0534^{***}	0.0216	0.0681^{*}	0.0820^{***}	0.0253^{**}
	(-0.0113)	(-0.0474)	(-0.0362)	(-0.0178)	(-0.0125)
Betweenness	0.0051	0.024	0.0232^{*}	0.013	0.00342
	(-0.00447)	(-0.024)	(-0.0135)	(-0.0083)	(-0.00582)
Transitivity	0.0088	0.00211	0.00685	-0.00705	0.0281^{**}
	(-0.00599)	(-0.0111)	(-0.0117)	(-0.013)	(-0.0127)
Structural H.	0.0145^{**}	0.0187	0.0227^{*}	0.00775	-0.00333
	(-0.00651)	(-0.012)	(-0.0135)	(-0.0132)	(-0.0132)
Eigenvector	-0.0163^{**}	0.194	-0.0498	0.0483^{*}	-0.00864
	(-0.00715)	(-0.124)	(-0.107)	(-0.0291)	(-0.00738)
Reach	-0.0759^{***}	-0.184^{***}	-0.163^{***}	-0.0416	-0.0549
	(-0.0236)	(-0.0535)	(-0.0524)	(-0.0494)	(-0.038)
Multinst	0.0385^{***}	0.0893^{*}	-0.0275	-0.0151	0.0676^{***}
	(-0.0142)	(-0.0493)	(-0.0351)	(-0.0318)	(-0.0191)
Interinst	-0.00978	-0.0397^{*}	0.0238	0.0175	-0.0278
	(-0.00984)	(-0.0237)	(-0.0211)	(-0.0209)	(-0.0176)
Multidisc	0.00946	0.0163	0.0347	0.0323	-0.0308^{*}
	(-0.0123)	(-0.0426)	(-0.0312)	(-0.0228)	(-0.0187)
Interdisc	-0.0114	-0.0233	-0.0241	-0.0480^{**}	0.0470^{**}
	(-0.0118)	(-0.0307)	(-0.0269)	(-0.023)	(-0.0222)
Gatekeeper	-0.00429	-0.0169	-0.0241^{*}	0.0194	0.00387
	(-0.00599)	(-0.0114)	(-0.0129)	(-0.0123)	(-0.0115)
International	0.0232	0.254^{**}	0.1	0.0569	0.00118
	(-0.0153)	(-0.126)	(-0.0625)	(-0.0371)	(-0.0131)
Year F.E.	Yes	Yes	Yes	Yes	Yes
Individual F.E.	Yes	Yes	Yes	Yes	Yes
Observations	48335	12082	12083	12079	12091
R-squared	0.042	0.01	0.027	0.058	0.119
Number of id_cvu	13727	4694	3313	2870	2850

reg	13727	4694	3313	2870	2850
betweenness	−0.002095	−0.011719	0.033466	0.022899	−0.016741
	(0.0197439)	(0.0509032)	(0.0455525)	(0.0139801)	(0.0145557)
closeness	0.034245	0.408456**	−0.206727	0.211291	0.304576*
	(0.1528771)	(0.149039)	(0.1945659)	(0.171488)	(0.1483526)
degree	0.017678	0.233826	0.133412	0.134946*	0.01479
	(0.0167949)	(0.1319713)	(0.0888846)	(0.0571796)	(0.0108368)
e_disc	0.018105	0.021891	−0.020089	−0.049505	0.007735
	(0.0364901)	(0.0353487)	(0.0780095)	(0.0316796)	(0.0302377)
e_inst	−0.017261	0.04214	−0.008359	0.027935	−0.059218**
	(0.0187151)	(0.0335756)	(0.0519703)	(0.0219806)	(0.0205854)
eigen_centrality	−0.003887	−3.884721	0.783721	0.02067	0.001001
	(0.0111184)	(4.839038)	(3.300965)	(0.7363972)	(0.0093179)
int_ratio	0.01401	−0.016376	−0.077268	−0.009492	0.041886**
	(0.0139904)	(0.0125314)	(0.0429909)	(0.0155958)	(0.0149723)
n_disc	$-2.1e-05$	−0.065069	0.01386	0.024067	−0.008426
	(0.0313335)	(0.0618087)	(0.0955075)	(0.0294512)	(0.0245662)
n_inst	0.069172**	−0.153182	−0.077213	−0.067391	0.064677***
	(0.0209641)	(0.0951892)	(0.106605)	(0.0353838)	(0.0159843)
n_int	−0.008607	−0.060436	0.319389	0.064292	−0.012474
	(0.0213717)	(0.2313912)	(0.6055442)	(0.0924251)	(0.0124503)
reach	−0.136836	−0.396386**	0.178375	−0.241424	−0.29585*
	(0.1393671)	(0.1389709)	(0.2069215)	(0.1654585)	(0.1429755)
struc_hole	−0.05568	0.101805**	0.048476	0.03145	−0.011666
	(0.0532246)	(0.038273)	(0.0428447)	(0.0471314)	(0.0554511)
transitivity	0.061358	−0.067869	−0.027937	0.013324	0.107647*
	(0.0540493)	(0.0417808)	(0.0409103)	(0.0491845)	(0.0505376)
st_age	−0.058778**	0.028926*	−0.042543	−0.038875**	−0.068288***
	(0.0172704)	(0.0112832)	(0.0266332)	(0.0122079)	(0.0126308)
st_pubs	0.063768**	0.036603	−0.045881	−0.065353*	0.003874
	(0.0236029)	(0.0653152)	(0.0764729)	(0.0268405)	(0.0181308)

a

	Dependent variable:		
	DWPP		
	Model 1	Model 2	Model 3
Citations	0.0534***	0.0534***	0.0119***
	(-0.00538)	(-0.00536)	(-0.00375)
Age	0.0632***	0.0632***	0.763***
	(-0.00382)	(-0.00382)	(-0.0544)
Closeness	−0.0729***	−0.0729***	0.0346*
	(-0.0178)	(-0.0178)	(-0.0188)
Degree	0.0411***	0.0411***	0.0243***
	(-0.00797)	(-0.00795)	(-0.00908)
Betweenness	0.0454***	0.0454***	0.0183***
	(-0.00522)	(-0.00522)	(-0.00379)
Transitivity	−0.0833***	−0.0833***	0.0149***
	(-0.00379)	(-0.00378)	(-0.00397)
Structural H.	−0.0473***	−0.0473***	0.000862
	(-0.00404)	(-0.00404)	(-0.00441)
Eigenvector	−0.00633	−0.00633	−0.00471
	(-0.00528)	(-0.00532)	(-0.00372)
Reach	0.0853***	0.0853***	−0.0336*
	(-0.0172)	(-0.0172)	(-0.0179)
Multinst	0.0509***	0.0509***	0.0332***
	(-0.0108)	(-0.0108)	(-0.00996)
Interinst	0.0122*	0.0122*	−0.0163**
	(-0.00647)	(-0.00646)	(-0.00691)
Multidisc	0.243***	0.243***	0.0632***
	(-0.00981)	(-0.00981)	(-0.00919)
Interdisc	−0.107***	−0.107***	−0.0634***
	(-0.00858)	(-0.00857)	(-0.00857)
Gatekeeper	0.0126***	0.0126***	−0.00279
	(-0.00347)	(-0.00347)	(-0.00399)
International	−0.0167**	−0.0167**	0.00512
	(-0.00751)	(-0.00752)	(-0.00805)
Year F.E.	No	Yes	Yes
Individual F.E.	No	No	Yes
Observations	48335	48335	48335
R-squared	0.219	0.219	0.047
Number of id_cvu			13727

7 Conclusion

a

	Dependent variable:						
	DWPP						
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Citations	0.0120***	0.0116***	0.00775	0.0148***	0.0121***	−0.0578	0.0119***
	(-0.00375)	(-0.00387)	(-0.0048)	(-0.00433)	(-0.00375)	(-0.162)	(-0.00376)
Age	0.771***	0.729***	0.753***	0.860***	0.744***	0.439	0.771***
	(-0.0543)	(-0.0629)	(-0.0554)	(-0.0899)	(-0.0561)	(-0.739)	(-0.0543)
Closeness	0.0138						
	(-0.00955)						
Degree		0.0931					
		(-0.075)					
Betweenness			0.216**				
			(-0.103)				
Transitivity				0.154			
				(-0.116)			
Structural H.					−0.0384		
					(-0.0237)		
Eigenvector						−3.822	
						(-9.646)	
Reach							0.0121
							(-0.00904)
Multinst	0.0550***	−0.0103	−0.00997	0.0647***	0.0521***	0.247	0.0553***
	(-0.00754)	(-0.0534)	(-0.0322)	(-0.0109)	(-0.00769)	(-0.462)	(-0.00755)
Interinst	−0.0255***	0.00075	−0.0185**	−0.0202***	−0.0319***	−0.184	−0.0252***
	(-0.00659)	(-0.0202)	(-0.00774)	(-0.00687)	(-0.00848)	(-0.382)	(-0.00656)
Multidisc	0.0734***	0.0691***	−0.0566	0.101***	0.0730***	−0.0186	0.0736***
	(-0.00901)	(-0.00987)	(-0.0633)	(-0.0224)	(-0.00901)	(-0.252)	(-0.00901)
Interdisc	−0.0731***	−0.0743***	0.00483	−0.0510***	−0.0794***	−0.0021	−0.0731***
	(-0.00849)	(-0.00873)	(-0.0381)	(-0.0181)	(-0.00961)	(-0.192)	(-0.0085)
Gatekeeper	−0.00155	−0.00192	0.00402	−0.00349	0.00276	−0.152	−0.00184
	(-0.0042)	(-0.00423)	(-0.00556)	(-0.00412)	(-0.00562)	(-0.345)	(-0.00417)
International	0.00596	−0.00909	0.0186	−0.00127	0.00541	1.89	0.00583
	(-0.00668)	(-0.0163)	(-0.0148)	(-0.00873)	(-0.00655)	(-4.337)	(-0.00667)
Year F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual F.E.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	48335	48335	48335	48335	48335	48335	48335
R-squared							
Number of id_cvu	13727	13727	13727	13727	13727	13727	13727

8 Discussion

This study used the framework of network theory to understand the effect of research collaboration on scientific performance. The network structure is a mechanism that models how resources, ideas, knowledge and skills are exchanged from social ties of RC. The analysis of the structure usually does not take into account node attributes and contextual elements of social ties that can have an important effect on the outcome. This study complement traditionally network structure analyses with a relational aspect of collaboration that captures the dimensions where RC happens and adds an additional layer of information about the content and diversity of information and resources flowing through peers. However, as has been pointed out, collaboration networks are conceived as a complex system "constituted by a variety of entities... that are largely autonomous, geographically distributed, and heterogeneous in terms of their: operating environment, culture, social capital,... that collaborate to better achieve common or compatible goals"(Camarinha-Matos & Afsarmanesh, 2005). This analysis derived conclusions from a singular dynamic network system composed only by the interactions of researchers and authors. A more comprehensive approach, that should motivate future analysis, is to disentangle the connections between different social arrangements from a multilayer system, or a network of networks (De Domenico, Lancichinetti, Arenas & Rosvall, 2015, 1). For instance, the connections between researchers have the social dimension, embedded in an institutional setting where knowledge is exchanged by disciplinary epistemic communities that interact

	Dependent variable:				
	DWPP				
	Model 1	Model 2	Model 3	Model 4	Model 5
Citations	0.0119***	0.00249	0.00283	−0.00404	0.0119**
	(-0.00375)	(-0.0169)	(-0.0113)	(-0.00884)	(-0.00464)
Age	0.763***	0.337***	0.737***	0.914***	1.215***
	(-0.0544)	(-0.0918)	(-0.105)	(-0.117)	(-0.106)
Closeness	0.0346*	0.0143	−0.0104	0.104***	−0.0111
	(-0.0188)	(-0.0349)	(-0.0423)	(-0.0398)	(-0.0306)
Degree	0.0243***	−0.0246	0.0422	0.0508***	0.0249*
	(-0.00908)	(-0.0317)	(-0.0281)	(-0.0135)	(-0.0134)
Betweenness	0.0183***	−0.0207	−0.0472***	0.0152	0.0104**
	(-0.00379)	(-0.0231)	(-0.0165)	(-0.0101)	(-0.00416)
Transitivity	0.0149***	0.0246***	0.0202**	0.0391***	−0.0210**
	(-0.00397)	(-0.00622)	(-0.00793)	(-0.00827)	(-0.0103)
Structural H.	0.000862	0.0182**	0.000803	−0.0052	−0.00749
	(-0.00441)	(-0.00711)	(-0.00874)	(-0.00905)	(-0.0108)
Eigenvector	−0.00471	−0.0603	−0.0489***	0.0225	−0.00311
	(-0.00372)	(-0.0993)	(-0.0161)	(-0.0156)	(-0.00394)
Reach	−0.0336*	−0.022	0.00986	−0.103***	0.0227
	(-0.0179)	(-0.0337)	(-0.0404)	(-0.0379)	(-0.0293)
Multinst	0.0332***	0.0318	0.0418*	−0.0131	0.0184
	(-0.00996)	(-0.0301)	(-0.0254)	(-0.0201)	(-0.0143)
Interinst	−0.0163**	−0.0322**	−0.0298**	0.0103	0.00968
	(-0.00691)	(-0.0144)	(-0.0144)	(-0.0149)	(-0.0133)
Multidisc	0.0632***	0.0324	−0.0161	−0.0191	0.0304**
	(-0.00919)	(-0.0337)	(-0.0287)	(-0.0231)	(-0.0124)
Interdisc	−0.0634***	−0.0601***	−0.0311	0.0121	−0.00638
	(-0.00857)	(-0.0222)	(-0.023)	(-0.0209)	(-0.0165)
Gatekeeper	−0.00279	−0.00066	−0.00803	−0.00797	$4.40E-05$
	(-0.00399)	(-0.00657)	(-0.00797)	(-0.0089)	(-0.00903)
International	0.00512	−0.0301	0.0233	0.0852***	0.00513
	(-0.00805)	(-0.0499)	(-0.0376)	(-0.0202)	(-0.00891)
Year F.E.	Yes	Yes	Yes	Yes	Yes
Individual F.E.	Yes	Yes	Yes	Yes	Yes
Observations	48335	12083	12082	12078	12092
R-squared	0.047	0.022	0.025	0.055	0.137
Number of id_cvu	13727	5212	3436	2762	2317

with multinational science systems.

	Dependent variable:				
	DWPP				
	Model 1	Model 2	Model 3	Model 4	Model 5
Citations	0.024724	−0.014863	−0.053341	−0.033158	0.002549
	(0.0235449)	(0.062314)	(0.0762663)	(0.0220289)	(0.0122833)
Age	0.095366***	0.013527	0.02453	0.028534	0.046678
	(0.0182947)	(0.0455793)	(0.0377933)	(0.0164095)	(0.0266915)
Closeness	0.087653	−0.0076	−0.076689	0.151919	0.244825
	(0.2370835)	(0.3139139)	(0.3198777)	(0.1650892)	(0.1546861)
Degree	0.01725	−0.248262	0.020501	0.049296	0.03753**
	(0.0196428)	(0.1445243)	(0.1044855)	(0.034669)	(0.0135345)
Betweenness	0.082361*	0.134026	−0.068437	0.05372	0.038948*
	(0.0338542)	(0.1222675)	(0.0545621)	(0.0354516)	(0.0177538)
Transitivity	−0.103721	0.075689*	0.018705	−0.006426	−0.011748
	(0.0566844)	(0.0297533)	(0.0436041)	(0.0411434)	(0.0599494)
Structural H.	0.00202	0.037875	0.003303	0.027269	0.051331
	(0.0553302)	(0.0409328)	(0.0384887)	(0.0410761)	(0.0565975)
Eigenvector	−0.016613	1.142134	8.372439	0.354037	−0.010355
	(0.0146174)	(2.334978)	(6.688477)	(0.7223745)	(0.013707)
Reach	−0.095507	0.007484	0.063311	−0.226789	−0.261569
	(0.2059594)	(0.3115089)	(0.285654)	(0.1586907)	(0.1442593)
Multinst	0.019694	0.080146	0.211519	−0.019242	0.002878
	(0.0327333)	(0.1494366)	(0.1755455)	(0.06187)	(0.017239)
Interinst	0.039517	−0.012258	−0.160744*	0.002142	−0.005904
	(0.0270801)	(0.0749711)	(0.0757093)	(0.0240135)	(0.0220878)
Multidisc	0.03344	−0.068993	−0.175676	−0.140094***	0.006392
	(0.0546824)	(0.1939678)	(0.1267858)	(0.0370895)	(0.0250848)
Interdisc	−0.068236	0.004289	0.027219	0.047634	−0.015726
	(0.0524895)	(0.1196929)	(0.0927709)	(0.0405348)	(0.0310282)
Gatekeeper	−0.030992	0.02796	0.06361	−0.025696	0.003893
	(0.0229989)	(0.0317489)	(0.0427284)	(0.0173369)	(0.017848)
International	0.052224*	0.046297	−1.434831	−0.057334	0.023682
	(0.0264077)	(0.7954489)	(0.9552197)	(0.1286013)	(0.0149503)

References

Abramo, G., Cicero, T. & D'Angelo, C. A. (2013). Individual research performance: a proposal for comparing apples to oranges. *Journal of Informetrics*, *7*(2), 528–539. doi:https://doi.org/10.1016/j.joi.2013.01.013

Adams, J. (2013). Collaborations: the fourth age of research. *Nature*, *497*(7451), 557–560.

Adler, P. S. & Kwon, S.-W. (2002). Social capital: prospects for a new concept. *The Academy of Management Review*, *27*(1), 17–40. Retrieved from http://www.jstor.org/stable/4134367

Ajiferuke, I., Burell, Q. & Tague, J. (1988, November). Collaborative coefficient: a single measure of the degree of collaboration in research. *Scientometrics*, *14*(5), 421–433. doi:10.1007/BF02017100

Amabile, T. M., Patterson, C., Mueller, J., Wojcik, T., Odomirok, P. W., Marsh, M. & Kramer, S. J. (2001). Academic-practitioner collaboration in management research: a case of cross-profession collaboration. *Academy of Management Journal*, *44*(2), 418–431. doi:10.2307/3069464. eprint: http://amj.aom.org/content/44/2/418.full.pdf+html

Amitai, G., Shemesh, A., Sitbon, E., Shklar, M., Netanely, D., Venger, I. & Pietrokovski, S. (2004). Network analysis of protein structures identifies functional residues. *Journal of Molecular Biology*, *344*(4), 1135–1146. doi:https://doi.org/10.1016/j.jmb.2004.10.055

Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: monte carlo evidence and an application to employment equations. *The Review of Economic Studies*, *58*(2), 277–297. doi:10.2307/2297968. eprint: /oup/backfile/content_public/journal/restud/58/2/10.2307/2297968/2/58-2-277.pdf

Berkun, S. (2010). *The myths of innovation*. O'Reilly Media. Retrieved from https://books.google.nl/books?id=kPCgnc70MSgC

Blume, L. E., Brock, W. A., Durlauf, S. N. & Ioannides, Y. M. (2011). Chapter 18 - identification of social interactions§§prepared for the handbook of social economics. In J. Benhabib, A. Bisin & M. O. Jackson (Eds.), (Vol. 1, pp. 853–964). Handbook of Social Economics. North-Holland. doi:https://doi.org/10.1016/B978-0-444-53707-2.00001-3

Blundell, R. & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, *87*(1), 115–143. doi:https://doi.org/10.1016/S0304-4076(98)00009-8

Blundell, R., Bond, S. & Windmeijer, F. (2001). Estimation in dynamic panel data models: improving on the performance of the standard gmm estimator. In *Nonstationary panels, panel cointegration, and dynamic panels* (Vol. Volume 15, pp. 53–91). doi:10.1016/S0731-9053(00)15003-0

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology*, *2*(1), 113–120. doi:10.1080/0022250X.1972.9989806. eprint: https://doi.org/10.1080/0022250X.1972.9989806

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71. doi:https://doi.org/10.1016/j.socnet.2004.11.008

Borgatti, S. P. & Halgin, D. S. (2011). On network theory. *Organization Science*, *22*(5), 1168–1181. doi:10.1287/orsc.1100.0641. eprint: https://doi.org/10.1287/orsc.1100.0641

Bozeman, B. & Corley, E. (2004). Scientists' collaboration strategies: implications for scientific and technical human capital. *Research Policy*, *33*(4), 599–616. Scientific and Technical Human Capital: Science Careers and Networks as Knowledge Assets. doi:https://doi.org/10.1016/j.respol.2004.01.008

Bramoullé, Y., Djebbari, H. & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, *150*(1), 41–55. doi:https://doi.org/10.1016/j.jeconom.2008.12.021

Buanes, A. & Jentoft, S. (2009). Building bridges: institutional perspectives on interdisciplinarity. *Futures*, *41*(7), 446–454. doi:https://doi.org/10.1016/j.futures.2009.01.010

Bukvova, H. (2010). Studying research collaboration: a literature review. *All Sprouts Content*. Retrieved from https://aisel.aisnet.org/sprouts_all/326

Bullmore, E. & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*, 186. Review Article. Retrieved from http://dx.doi.org/10.1038/nrn2575

Burt, R. (2001). Structural holes versus network closure as social capital. In R. Dubos (Ed.), *The oxford handbook of innovation* (Chap. 2). New York: Routledge.

Camarinha-Matos, L. M. & Afsarmanesh, H. (2005, October). Collaborative networks: a new scientific discipline. *Journal of Intelligent Manufacturing*, *16*(4), 439–452. doi:10.1007/s10845-005-1656-3

Cowan, R. & Jonard, N. (2004). Network structure and the diffusion of knowledge. *Journal of Economic Dynamics and Control*, *28*(8), 1557–1575. doi:https://doi.org/10.1016/j.jedc.2003.04.002

Crespi, G. A. & Geuna, A. (2008). An empirical study of scientific production: a cross country analysis, 1981–2002. *Research Policy*, *37*(4), 565–579. doi:https://doi.org/10.1016/j.respol.2007.12.007

De Carolis, D. M. & Saparito, P. (2006). Social capital, cognition, and entrepreneurial opportunities: a theoretical framework. *Entrepreneurship Theory and Practice*, *30*(1), 41–56. doi:10.1111/j.1540-6520.2006.00109.x

De Domenico, M., Lancichinetti, A., Arenas, A. & Rosvall, M. (2015, March). Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems. *Phys. Rev. X*, *5*, 011027. doi:10.1103/PhysRevX.5.011027

Ductor, L. (2015). Does co-authorship lead to higher academic productivity? *Oxford Bulletin of Economics and Statistics*, *77*(3), 385–407. doi:10.1111/obes.12070

Ductor, L., Fafchamps, M., Goyal, S. & van der Leij, M. J. (2014). Social networks and research output. *The Review of Economics and Statistics*, *96*(5), 936–948. doi:10.1162/REST_a_00430. eprint: https://doi.org/10.1162/REST_a_00430

Fafchamps, M., van der Leij, M. J. & Goyal, S. (2010). Matching and network effects. *Journal of the European Economic Association*, *8*(1), 203–231. Retrieved from http://www.jstor.org/stable/40601219

Freeman, L. (1978, January). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. doi:10.1016/0378-8733(78)90021-7

Freeman, R. B. & Huang, W. (2014). Collaboration: strength in diversity. *Nature*, *513*(7518), 305. 305.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P. & Trow, M. (1994). *The new production of knowledge: the dynamics of science and research in contemporary societies*. SAGE Publications.

Glänzel, W. & Schubert, A. (2005). Analysing scientific networks through co-authorship. In H. F. Moed, W. Glänzel & U. Schmoch (Eds.), *Handbook of quantitative science and technology research: the use of publication and patent statistics in studies of s&t systems* (pp. 257–276). Dordrecht: Springer Netherlands. doi:10.1007/1-4020-2755-9_12

Goldsmith-Pinkham, P. & Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, *31*(3), 253–264. doi:10.1080/07350015.2013.801251. eprint: http://dx.doi.org/10.1080/07350015.2013.801251

Gonzalez-Brambila, C. N. [Claudia N.], Veloso, F. M. & Krackhardt, D. (2013). The impact of network embeddedness on research output. *Research Policy*, *42*(9), 1555–1567. doi:10.1016/j.respol.2013.07.008

Gonzalez-Brambila & Veloso, F. M. (2007). The determinants of research output and impact: a study of mexican researchers. *Research Policy*, *36*(7), 1035–1051. doi:https://doi.org/10.1016/j.respol.2007.03.005

Gould, R. V. & Fernandez, R. M. (1989). Structures of mediation: a formal approach to brokerage in transaction networks. *Sociological Methodology*, *19*, 89–126. Retrieved from http://www.jstor.org/stable/270949

Haas, P. M. (1992). Introduction: epistemic communities and international policy coordination. *International Organization*, *46*(1), 1–35. doi:10.1017/S0020818300001442

Hara, N., Solomon, P., Kim, S.-L. & Sonnenwald, D. H. (2003). An emerging view of scientific collaboration: scientists' perspectives on collaboration and factors that impact collaboration. *Journal of the American Society for Information Science and Technology*, *54*(10), 952–965. doi:10.1002/asi.10291

Haythornthwaite, C. (1996). Social network analysis: an approach and technique for the study of information exchange. *Library Information Science Research*, *18*(4), 323–342. doi:https://doi.org/10.1016/S0740-8188(96)90003-1

Hicks, D. (2005). The four literatures of social science. In H. F. Moed, W. Glänzel & U. Schmoch (Eds.), (pp. 473–496). Dordrecht: Springer Netherlands. doi:10.1007/1-4020-2755-9_22

Hicks, D. M. & Katz, J. S. (1996). Where is science going? *Science, Technology, Human Values*, *21*(4), 379–406. Retrieved from http://www.jstor.org/stable/690087

Horrace, W. C., Liu, X. & Patacchini, E. (2016). Endogenous network production functions with selectivity. *Journal of Econometrics*, *190*(2), 222–232. Endogeneity Problems in Econometrics. doi:https://doi.org/10.1016/j.jeconom.2015.06.005

Hsieh, C.-S. & Lee, L. f. (2017, April). Specification and Estimation of Network Formation and Network Interaction Models with the Exponential Probability Distribution. (60726). Retrieved from https://ideas.repec.org/p/pra/mprapa/60726.html

Huang, S., Yang, B., Yan, S. & Rousseau, R. (2014, June). Institution name disambiguation for research assessment. *Scientometrics*, *99*(3), 823–838. doi:10.1007/s11192-013-1214-2

Jassawalla, A. R. & Sashittal, H. C. (1998). An examination of collaboration in high-technology new product development processes. *Journal of Product Innovation Management*, *15*(3), 237–254. doi:https://doi.org/10.1016/S0737-6782(97)00080-5

Katz, J. & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18. doi:https://doi.org/10.1016/S0048-7333(96)00917-1

Kramer, R. (1990). Reviewed work: collaborating: finding common ground for multiparty problems. *The Academy of Management Review*, *15*(3), 545–547. Retrieved from http://www.jstor.org/stable/258026

Kumar, S. (2015). Co-authorship networks: a review of the literature. *Aslib Journal of Information Management*, *67*(1), 55–73. doi:10.1108/AJIM-09-2014-0116

Leydesdorff, L. & Wagner, C. S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics*, *2*(4), 317–325. doi:https://doi.org/10.1016/j.joi.2008.07.003

Liu, Y., Rafols, I. & Rousseau, R. (2012). A framework for knowledge integration and diffusion. *Journal of Documentation*, *68*(1), 31–44. doi:10.1108/00220411211200310. eprint: https://doi.org/10.1108/00220411211200310

Marsden, P. V. (2002). Egocentric and sociocentric measures of network centrality. *Social Networks*, *24*(4), 407–422. doi:https://doi.org/10.1016/S0378-8733(02)00016-3

Melin, G. & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, *36*(3), 363–377. doi:10.1007/BF02129600. eprint: http://akademiai.com/doi/pdf/10.1007/BF02129600

Melis, A. P. (2013). The evolutionary roots of human collaboration: coordination and sharing of resources. *Annals of the New York Academy of Sciences*, *1299*(1), 68–76. doi:10.1111/nyas.12263

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, *98*(2), 404–409. doi:10.1073/pnas.98.2.404. eprint: http://www.pnas.org/content/98/2/404.full.pdf

Nissani, M. (1997). Ten cheers for interdisciplinarity: the case for interdisciplinary knowledge and research. *The Social Science Journal*, *34*(2), 201–216. doi:https://doi.org/10.1016/S0362-3319(97)90051-3

Rafols, I. & Meyer, M. (2010, February). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, *82*(2), 263–287. doi:10.1007/s11192-009-0041-y

Rodriguez Miramontes, J. & Gonzalez-Brambila, C. N. [C. N.]. (2016, November). The effects of external collaboration on research output in engineering. *Scientometrics*, *109*(2), 661–675. doi:10.1007/s11192-016-2054-7

Rørstad, K. & Aksnes, D. W. (2015). Publication rate expressed by age, gender and academic position – a large-scale analysis of norwegian academic staff. *Journal of Informetrics*, *9*(2), 317–333. doi:https://doi.org/10.1016/j.joi.2015.02.003

Sabidussi, G. (1966, December). The centrality index of a graph. *Psychometrika*, *31*(4), 581–603. doi:10.1007/BF02289527

Schulz, C., Mazloumian, A., Petersen, A. M., Penner, O. & Helbing, D. (2014, September). Exploiting citation networks for large-scale author name disambiguation. *EPJ Data Science*, *3*(1), 11. doi:10.1140/epjds/s13688-014-0011-3

Simonton, D. (2004). *Creativity in science: chance, logic, genius, and zeitgeist.* Cambridge University Press. Retrieved from https://books.google.nl/books?id=00KsMKjXxzYC

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, *41*(1), 643–681. doi:10.1002/aris.2007.1440410121

Spizzirri, L. (2011). Justification and application of eigenvector-centrality. http://blog.narotama.ac.id/wp-content/uploads/2015/02/Justification-and-Application-of-Eigenvector-Centrality.pdf. (Accessed on 03/29/2018).

Stephan, P. E. (1996). The economics of science. *Journal of Economic Literature*, *34*(3), 1199–1235. Retrieved from http://www.jstor.org/stable/2729500

Tichy, N. M., Tushman, M. L. & Fombrun, C. (1979). Social network analysis for organizations. *Academy of Management Review*, *4*(4), 507–519. doi:10.5465/AMR.1979.4498309. eprint: http://amr.aom.org/content/4/4/507.full.pdf+html

Valente, T. W., Coronges, K., Lakon, C. & Costenbader, E. (2008, January). How correlated are network centrality measures? *Connect (Tor)*, *28*(1), 16–26. 20505784[pmid]. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2875682/

Wagner, C. S. (2005). Six case studies of international collaboration in science. *Scientometrics*, *62*(1), 3–26. doi:10.1007/s11192-005-0001-0. eprint: http://akademiai.com/doi/pdf/10.1007/s11192-005-0001-0

Wood, D. J. & Gray, B. (1991). Toward a comprehensive theory of collaboration. *The Journal of Applied Behavioral Science*, *27*(2), 139–162. doi:10.1177/0021886391272001. eprint: https://doi.org/10.1177/0021886391272001

Wuchty, S., Jones, B. F. & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039. doi:10.1126/science.1136099. eprint: http://science.sciencemag.org/content/316/5827/1036.full.pdf

Yamamoto, K. & Ishikawa, M. (2017). A normalized index to quantify publication productivity across disciplinary boundaries. *International Journal of Institutional Research and Management*, *1*(1), 67–81. Retrieved from http://www.iaiai.org/journals/index.php/IJIRM/article/view/114

Yan, E. & Ding, Y. (2009). Applying centrality measures to impact analysis: a coauthorship network analysis. *Journal of the American Society for Information Science and Technology*, *60*(10), 2107–2118. doi:10.1002/asi.21128