

8. The relationship between the research performance of scientists and their position in co-authorship networks in three fields

Artículo publicado en Journal of Informetrics, 9, 135-144. doi: [10.1016/j.joi.2014.12.001](https://doi.org/10.1016/j.joi.2014.12.001). Autores: María Bordons, Javier Aparicio, Borja González-Albo, Adrián A. Díaz-Faes.

Abstract: Research networks play a crucial role in the production of new knowledge since collaboration contributes to determine the cognitive and social structure of scientific fields and has a positive influence on research. This paper analyses the structure of co-authorship networks in three different fields (Nanoscience, Pharmacology and Statistics) in Spain over a three-year period (2006–2008) and explores the relationship between the research performance of scientists and their position in co-authorship networks. A denser co-authorship network is found in the two experimental fields than in Statistics, where the network is of a less connected and more fragmented nature. Using the g-index as a proxy for individual research performance, a Poisson regression model is used to explore how performance is related to different co-authorship network measures and to disclose interfield differences. The number of co-authors (degree centrality) and the strength of links show a positive relationship with the g-index in the three fields. Local cohesion presents a negative relationship with the g-index in the two experimental fields, where open networks and the diversity of co-authors seem to be beneficial. No clear advantages from intermediary positions (high betweenness) or from being linked to well-connected authors (high eigenvector) can be inferred from this analysis. In terms of g-index, the benefits derived by authors from their position in co-authorship networks are larger in the two experimental fields than in the theoretical one.

Keywords: Research performance, Collaboration, Social network analysis, Co-authorship, G-index, Poisson regression model.

8.1. Introduction

Science is increasingly becoming a collaborative endeavour. Collaboration allows scientists to share knowledge, expertise and techniques, expedites the research process, and increases visibility (Katz & Martin, 1997 and Sonnenwald, 2007). Under the assumption of the importance and benefits of collaboration for the advancement of science, scientific collaboration is encouraged by policy makers and the collaboration process is the subject of many academic studies.

From a bibliometric standpoint, collaboration is usually analysed through co-authorship in scientific publications. This indicator presents several limitations, since all co-authorships are sometimes not based on collaborative contributions (e.g. honorary authorship) and not all authors who collaborate become co-authors (Laudel, 2002). However, a positive correlation between collaboration and co-authorship has been described in the literature and this indicator has proved useful to study different aspects of collaboration in science (see for example, Glänzel & Schubert, 2004). Accordingly, co-authorship is used as a measure of scientific collaboration in this paper, although we should have in mind its limitations.

Different indicators have been introduced to quantify collaboration in research papers (see for example, Egghe, 1991; Glänzel & Schubert, 2004 and Vinkler, 2010) and extensive literature has been devoted to explore collaboration patterns (Bordons & Gómez, 2000) and the influence of collaboration on the productivity of scientists and on the impact of research (Abramo, D'Angelo & Di Costas, 2009; Bordons, Aparicio & Costas, 2013; Glänzel, 2001 and Lee & Bozeman, 2005). In most recent years, the application of social network analysis to study co-authorship relations has emerged as an interesting approach, since it allows us to visualise and investigate social structures and relations (see for example, Abbasi, Altmann, & Hossain, 2011; Abbasi, Chung, & Hossain, 2012; Jansen, Von Görtz, & Heidler, 2010; Li-Chun, Kretschmer, Hanneman, & Ze-Yuan, 2006; Newman, 2001 and Otte & Rousseau, 2002). Studies of co-authorship networks may focus on the global structure of networks (macro-perspective) (see for example, Newman, 2001), on the study of subsets (clusters or components) formed within the network (meso-perspective) (He, Ding, & Ni, 2011) or on the individual scientists included in the network's membership (micro-perspective) (for example, Hou, Kretschmer, & Liu, 2008).

Different studies suggest that research networks play a crucial role in the production of new knowledge. The basic idea is that “the position of a node in a network determines in part the opportunities and constraints that it encounters, and in this way plays an important role in a node's outcomes” (Borgatti, Mehra, Brass, & Labianca, 2009). In other words, this means that the position of a scientist in the co-authorship network may have an influence on his/her research performance. This is clearly related to the notion of “social capital”, defined as the benefits that actors derive from their social relationships (Coleman, 1988), which may contribute to knowledge creation and to human capital development (Liao, 2011). Three different dimensions of social capital have been described (Nahapiet & Ghoshal, 1998), namely, cognitive capital, relational capital, and structural capital. The latter is the main subject-matter of this study and it can be defined as the value or advantage accrued by an individual or group arising from the structure of social relationships.

There is no consensus on which type of network structure performs best. According to Coleman (1988), densely embedded closed networks are advantageous because they foster the building of mutual confidence and partners bind themselves to one another through reciprocal obligations and expectations ("closure argument"). On the other hand, an alternative view considers that social structural advantages derive from the brokerage opportunities created by an open social structure (Burt, 1992 and Burt, 2004), since it fosters the flow of knowledge between heterogeneous actors and reduces redundant contacts. From this perspective, separate groups control different information and resources, and individuals who bring together people from the different groups act as "brokers" that bridge the existing gaps or "structural holes"²³ between groups ("structural hole argument"). Interestingly, these two notions of social capital are not necessarily contradictory, since different network structures may generate social capital depending on the purpose of the network and the members involved (Ahuja, 2000 and Klenk, Hickey, & MacLellan, 2010).

The relationship between the position of authors in collaboration networks and their performance, as measured by the number of publications, the number of citations and/or the h-index or the g-index, as the case may be, has been previously analysed in the literature. A positive correlation between different centrality measures and citation counts has been described in the fields of information systems (Liao, 2011) and library and information science (Yan & Ding, 2009), while centrality measures showed a positive correlation with scientific output in scientometrics (Hou et al., 2008); these results suggesting that researchers with a higher number of collaborators (high degree) or those who are close to all others in the network (high closeness) are likely to obtain better performance results. Moreover, the influence of the strength of the ties among authors has attracted considerable attention in a number of studies. Scholars who have strong ties (repeated co-authorships) to co-authors earned better research performance results than those with weak ties (single co-authorships with many different authors) in a study on information science (Abbasi et al., 2011). In this study, having an efficient network, with a low rate of redundant contacts, enhanced research performance probably because redundant contacts are less frequently associated with groundbreaking initiatives since they do not provide access to new information. Conversely, establishing connections with researchers in new and diverse teams, bridging structural holes, appeared to be positive for research performance. A positive effect of structural holes on a researcher's performance, as measured by citation scores and individual creativity, was described also in a study in nanoscience (Heinze & Bauer, 2007), while the development of closed social networks with strong ties was positive in other studies on the biotechnology (Walker, Kogut, & Shan, 1997) and pharmaceutical industries (Guler & Nerkar, 2012). As mentioned above, the effect

²³ A structural hole is the absence of ties among a pair of nodes in the ego network (Burt, 1992). The ego is the individual, team or organisational unit under analysis.

of structural holes on performance may vary depending on the context and the field. In this sense, Ahuja (2000) suggests that closed networks are beneficial when strong collaboration is required, while structural holes are likely to be more advantageous when access to diverse information is essential. On the other hand, the positive effect of structural holes may be higher in new fields (such as nanoscience) where brokerage positions become particularly significant because diverse knowledge and ideas are essential for the development of the field.

The patterns and consequences of network structures on scientific or innovative results have been studied in the literature at different levels of analysis, which range from individual scientists (Hou et al., 2008, Klenk et al., 2010 and Li-Chun et al., 2006) or teams (Reagans & Zuckerman, 2001), to higher organisational units such as firms (Ahuja, 2000 and Guler & Nerkar, 2012). Most of these studies deal with the analysis of publications on a given topic or field, whereas interfield comparisons are less frequently addressed. Special mention must be made of the study by Jansen et al. (2010) on the fields of astrophysics and nanoscience concluding that the relationship between network structure and the production of new knowledge is field specific, probably because fields differ in their cognitive structure and knowledge production dynamics.

The objective of this paper is to study the co-authorship networks existing in three different fields (macro-perspective) and explore the relationship between social network measures and research performance of authors with special emphasis on interfield differences (micro-perspective). The assumption that the experimental/theoretical character of a field and its degree of interdisciplinarity may have an influence on its cognitive and social structure was used as the main driver for the selection of our fields of study: one experimental and well-established field (Pharmacology), one experimental, emergent and interdisciplinary field (Nanoscience), and a theoretical field (Statistics).

The interest of this type of study is manifold. The analysis of the fields' structure through the study of their co-authorship networks and the examination of the relationship between social network measures and the research performance of authors may enable us to understand knowledge production dynamics in each field, to figure out which practices are linked to higher performance results and to identify the authors who have a more strategic position within the networks.

8.2. Research question

The questions addressed in this study are as follows:

- ✓ What are the main differences in the structure of fields according to social network measures based in co-authorship analysis?
- ✓ Is there any relationship between the position of a scientist within his/her co-authorship network and his/her research performance? If so, which of the social network-based measures shows a stronger relationship with the performance of scientists? Are there any interfield differences?

8.3. Methods

Scientific publications of Spain on Statistics/Probability, Pharmacy/Pharmacology and Nanoscience/Nanotechnology over the 2006-2008 period were downloaded from the Web of Science database (Science Citation Index Expanded, Social Science Citation Index and Arts & Humanities Citation Index). Disciplines were defined according to the classification of journals into subfields described by the Web of Science.

To cope with the inconsistencies in the names of authors, we used different algorithms aimed at the normalisation of names. These take into account text similarity of names, number of collaborators in common, number of publication journals in common and author subfield to identify pairs of names that are likely to correspond to the same author (Costas & Bordons, 2007).

A matrix including co-authorship frequencies was built for the social network analysis, while different research performance measures were calculated for individual scientists.

8.3.1. Social network measures

Social networks are usually represented by graphs, which include nodes and links. In this paper, nodes correspond to authors and links represent the cooperation relationship between authors on a joint publication. The Pajek software (Batagelj & Mrvar, 2013) was used to graph the network (not shown in this paper) and to calculate the network measures, which can be grouped in two different types: (a) centrality measures and (b) measures of cohesion.

Centrality measures

Centrality measures are useful to analyse how “central” an individual node is to a network. Different measures have been described:

- ✓ *Degree centrality*: This is the number of other nodes connected directly to a given node; therefore, in a co-authorship network, the degree of an author is the number of his/her different co-authors. It is a measure of local centrality (Scott,

1991). Since our interest was to compare node centrality across fields, which have networks with different sizes, a standardised value was calculated. The standardised degree centrality normalises the actual number of links by the maximum number of links it could have (Freeman, 1979), that is: $\text{normalised degree} = \text{degree} / (n - 1)$, where n is the number of nodes in the network. The normalised degree ranges from 0 (isolated node) to 1 (if the node is connected to all others).

- ✓ Closeness centrality: A node is globally central if it lies in average at the shortest distance from all other nodes. It focuses on “how close” an actor is to all other actors in the network (Freeman, 1979). Degree centrality identifies actors who are locally influential (it takes into account the immediate links that a node has), but closeness centrality focuses on the influence of a node over the entire network. A standardised value was calculated to make inter-field comparisons possible: $\text{normalised closeness} = \text{closeness} / (n - 1)$. The normalised closeness ranges from 0 to 1. This index is only meaningful for a connected network, so it was only applied to the main component.
- ✓ Betweenness centrality: The betweenness centrality of a vertex in a graph is calculated as the number of geodesics passing through that vertex. A geodesic is the shortest path between two vertices. In a connected, undirected graph with n vertices, there are at least $n(n - 1)$ geodesics. The betweenness centrality can be normalised using $(n - 1)(n - 2)/2$, which is the maximum number of shortest paths (excluding the node under consideration) (Abbasi et al., 2011). The normalised $\text{betweenness} = \text{betweenness} / [(n - 1)(n - 2)/2] = 2 * \text{betweenness} / (n^2 - 3n + 2)$. It ranges from 0 (a node lies on all geodesics of all pairs of nodes) to 1 (a node lies on no geodesic). In social networks, actors with high betweenness represent gatekeepers or information brokers because they lie among many paths of information flow.
- ✓ Centralisation: Centrality measures characterise an actor's position in a network (micro-level measure), while centralisation characterises the whole network (macro-level measure). It indicates how unequal the distribution of centrality is in a network. Degree centralisation in a network is calculated as the variation in the degrees of vertices divided by the maximum degree which is possible in the network of the same size (Wasserman & Faust, 1994). Networks where one or a few nodes show much higher centrality than the other nodes are highly centralised while those in which centrality measures do not differ significantly among nodes show low centralisation. It ranges from 0 (low centralisation) to 1 (high centralisation). In the same way, the betweenness centralisation was calculated.

Closeness centralisation is not shown because it is meaningful only for connected networks.

- ✓ Eigenvector centrality: This takes into account not only the number of adjacent nodes but also the values of centrality of these adjacent nodes assuming that a node which is connected to many other nodes that are themselves well-connected has a high eigenvector centrality. Kleimberg (1999) method is used, that is close to Bonacich power (Bonacich, 1972).

Measures of cohesion

Various measures related to the structural cohesion of the networks were considered.

- ✓ Strength of ties: The strength of a tie between node i and j is the weight of the link w_{ij} between those nodes. The weight is the number of co-authorships between two scholars. To assess a node ties strength we obtained the average of the weights of his co-authorships, that is, the number of co-authorships divided by the node degree (Abbasi et al., 2011).
- ✓ Network constraint: This allows assessing whether the research networks of the research groups are concentrated directly or indirectly on a single contact (which means no access to structural holes) (Burt, 1992), that is, it allows us to measure how open or closed research networks are. It can be calculated as follows:

$$C_{ij} = (p_{ij} + \sum_q p_{iq} p_{qj})^2, \text{ for } q \neq i, j.$$

P_{ij} is the proportion of i 's relations directly invested in connection with j . The next figure in brackets is the proportion of i 's relations that are indirectly invested in connection with contact j (Burt, 2004). Constraint is a measure of redundancy of contacts. If an individual's contacts are highly connected to each other, he/she has many redundant contacts and his/her network is highly constrained (Abbasi et al., 2012).

- ✓ Clustering coefficient: This is the average of the densities of the neighbourhoods of all actors in a network. It measures to what extent each actor in a network is "embedded" in a local cluster. It is the probability that two neighbours of a vertex are adjacent to each other, that is, the probability that two of a scientist's collaborator have themselves collaborated (Abbasi et al., 2011 and Barabási et al., 2002). A low clustering coefficient for an author means that his/her non-connected co-authors have low probability of writing a joint paper. This is the measure which

provides more specific information about cohesion. When this measure in a network is high, all actors are embedded in cohesive local neighbourhoods (Hanneman & Riddle, 2005).

8.3.2. Measures of research performance

The following indicators were calculated for each author.

- ✓ Number of articles: number of articles published in journals covered by the Web of Science database (WoS) (Science Citation Index Expanded, Social Sciences Citation Index, and Arts & Humanities Citation Index). Only articles, reviews and proceedings papers were considered.
- ✓ Total number of citations received by the articles in WoS journals. Citations from publication year to February 2014 were counted.
- ✓ Number of citations per article: This is the average number of citations received by articles published by a given scientist.
- ✓ G-index: Given a set of articles ranked in descending order of the number of citations received, the g-index is the (unique) highest number so that the top g articles received (altogether) at least g^2 citations (Egghe, 2006). The advantage of the g-index is that it measures quantity and impact of research by means of a single indicator. The g-index was introduced in 2006 as an improvement to Hirsch's h-index (Hirsch, 2005) because it takes into account the citation scores of top articles and this yields a more precise distinction between scientists from the point of view of visibility.

A regression analysis was used to explore to what extent there is a relationship between the g-index of scientists and their position in the social networks (social network measures as explanatory variables). Since the g-index tends to approximate the form of a Poisson distribution (it takes only positive integer values, it exhibits a positive skew, and the mean and the variance show very similar results), the Poisson multiple regression model was retained. In addition, the Kruskal Wallis test was applied to compare g-index distribution (which does not comply with the normal distribution assumption) between multiple groups of authors. The α level was fixed at 5%. Statistical analyses were conducted with SPSS (version 19)²⁴.

²⁴ As an alternative regression model, the relationship between the variable No. Publications*No. Citations (dependent variable) and social network based measures (explanatory variables) was also explored using a negative binomial regression. The results obtained were very similar to those derived from the g-index based regression, but lower R^2 values were achieved, so only the first model is shown.

8.4. Results

The scientific output of Spain for the 2006–2008 period amounted to 943 articles in Statistics/Probability; 1087, in Nanoscience/Nanotechnology; and 2858 in Pharmacology/Pharmacy. The total number of authors in each field and the resulting productivity per author are shown in the first panel of Table 8.1. Higher average team size is observed in Nanoscience and Pharmacology than in Statistics in accordance with their higher co-authorship index (CI), which can be accounted for by a stronger need for collaboration in the experimental fields. An in-depth study of co-authorship links is further conducted through social network analysis.

Table 8.1. General description of the networks (macro-level).

	Statistics	Nanoscience	Pharmacology
Total network			
No.Articles	943	1,087	2,858
No. Authors	1,572	3,505	10,099
N.Art./author	0.60	0.31	0.28
No.Authors/art (CI)	1.67	3.22	3.53
Reduced network*			
No.Authors	429	1,013	2,609
No.Edges	603	3,106	9,410
Degree centralisation	0.033	0.038	0.025
Betweenness centralisation	0.049	0.165	0.044
No.Components	80	75	162
No.Authors in main component (%)	119 (27.74%)	609 (60.12%)	1,731 (66.35%)
Mean distance	5.34	7.63	8.04
Largest distance	14	23	22

*Reduced network: only non-isolated authors (degree>0) with more than 1 article are considered.

8.4.1. Network structure

A general description of the networks is shown, first at the macro-level to depict the structure of the entire network (second panel of Table 8.1), and then at the micro-level through different measures that characterise the behaviour of authors on the basis of their relationships with other authors (Table 8.2). This study focuses on the set of non-isolated authors (degree > 0) with more than 1 article which constitutes what we have termed as the “reduced network” and is the subject of study in this research paper.

Table 8.2. Structural network measures of authors (micro-level).

	Statistics (n=429)		Nanoscience (n=1,013)		Pharmacology (n=2,609)	
	Av	SD	Av	SD	Av	SD
Degree	2.81	1.90	6.13	4.51	7.21	5.52
Std_Degree	.007	.004	.006	.004	.003	.002
Std_Closeness	.021	.020	.052	.039	.058	.041
Std_Betweenness	.001	.004	.002	.009	.001	.004
Eigenvector	.005	.048	.004	.031	.001	.020
Clustering coefficient	.611	.431	.776	.300	.775	.277
Constraint	.843	.246	.596	.256	.542	.255
Strength	2.10	0.88	1.85	0.57	1.83	0.76

Note: Av = average; SD = Standard deviation; Std = standardised.

Reduced network: only non-isolated authors (degree>0) with more than 1 article are considered.

There are several differences among the three fields worth pointing out. Firstly, a dense network is observed in Nanoscience and Pharmacology, where the number of lines is far higher than the number of vertices, whereas the network in Statistics may be qualified as sparse (De Nooy, Mrvar, & Batagelj, 2005) since the number of lines in the graph is of the same order as the number of vertices. Secondly, the networks show low values of centralisation in all three fields, that is, centrality is not concentrated in a low number of nodes. Anyway, Nanoscience, if any, is the field which shows the highest centralisation, especially concerning betweenness, because a few authors show relatively high betweenness values. Thirdly, the main component includes around two thirds of the authors in the denser networks (Pharmacology and Nanoscience), as against only 28% in Statistics, which shows a more fragmented structure. One of the underlying reasons for this divergence rests with the fact that collaboration is essential in experimental fields, such as Pharmacology and Nanoscience, where laboratory teamwork is essential; while it is not so indispensable in theoretical fields such as Statistics, where scientists are more likely to work alone or in small teams.

Table 8.2 shows the summary statistics of the structural network measures of authors in the three fields under analysis. Pharmacology and Nanoscience are quite similar according to the patterns of relationships of their authors, while Statistics shows a closer network (higher constraint), stronger links between authors (higher strength) and weaker local cohesion according to the lower propensity of authors to form cliques (lower clustering coefficient). The high number of articles with 2–3 authors in Statistics (68% vs. 20% in Pharmacology and Nanotechnology) contributes to explain its

higher constraint, since authors with a high number of collaborators (high degree) are more likely to have non-redundant contacts. On the other hand, the high number of authors with only one co-author in Statistics (22% of authors vs. 4–6% in Pharmacology and Nanoscience, respectively) contributes to explain the lower propensity of authors to form cliques in that field, since at least two co-authors are needed to form a clique²⁵.

Summary statistics of the performance of authors by field are shown in Table 8.3. Although outstanding interfield differences in the average citedness of authors' papers or in the g-index of scientists are observed, they cannot be compared due to differences in publication and citation practices by field (Moed, 2005).

Table 8.3. Summary statistics of the performance of authors by field.

	Statistics	Nanoscience	Pharmacology
N.Authors	429	1,013	2,609
N.Art./author	3.49 (2.6)	3.11 (2.01)	3.29 (3.05)
No.Cit/art.	1.38 (2.49)	21.34 (23.38)	18.86 (17.95)
G-index	1.21 (1.02)	2.96 (1.78)	3.08 (2.10)

8.4.2. Relationship between performance indicators and the position of authors in networks

To explore to what extent changes in the co-authorship network measures contribute to explain changes in the g-index we used a Poisson regression model. The predictor variables introduced include seven continuous ones for each of the fields: standardised betweenness, standardised closeness, standardised degree, eigenvector, clustering coefficient, average ties strength and constraint. Unfortunately, constraint had to be removed from the analysis due to multicollinearity problems. To allow for the comparison of variables which are expressed in different units of measurement, continuous variables are transformed to new variables with a mean of 0 and a standard deviation of 1 (Z-scores). Z-scores are a unit free measure which can be used to compare observations measured with different units. Three different models are built, one for each field, to identify interfield differences in the association between the co-authorship measures and the g-index.

²⁵ If degree = 1, the clustering coefficient = 0, since no cliques can be formed.

Our results show that the models fit reasonably well. The omnibus test, which compares the fitted model against the intercept-only model, is statistically significant in all three fields ($p < 0.001$) suggesting that changes in the predictor variables contribute to explain changes in the dependent variable. The results of the Poisson regression model are shown in Table 8.4. As a measure of the goodness of fit of the models the correlations between observed and predicted values of g-index are calculated. The best fit is obtained in Pharmacology ($R^2 = 0.652$), followed by Nanoscience ($R^2 = 0.573$) and Statistics ($R^2 = 0.195$).

Table 8.4. Poisson regression analysis for the g-index.

	Statistics				Nanoscience				Pharmacology			
	θ	Hypothesis test		Exp(θ)	θ	Hypothesis test		Exp(θ)	θ	Hypothesis test		Exp(θ)
		Wald Chi Square	Sig.			Wald Chi Square	Sig.			Wald Chi Square	Sig.	
(Intercept)	.147	10.38	.001	1.159	1.014	2,744.62	0.000	2.756	1.042	7,713.91	0.000	2.835
ZStd_degree	.277	32.11	.000	1.319	.200	103.19	0.000	1.221	.157	256.84	0.000	1.170
Zcloseness	-.092	2.75	.097	.912	.032	1.94	.164	1.032	.041	10.03	.002	1.042
Zbetweenness	-.050	1.20	.274	.951	-.021	2.25	.134	.979	.013	1.69	.193	1.013
Zclust_coefficient	-.069	2.15	.143	.933	-.167	85.17	0.000	.846	-.169	204.24	0.000	.844
Zav_strength	.192	23.60	.000	1.211	.198	123.44	0.000	1.219	.255	939.69	0.000	1.291
Zeigenvector	-.022	0.45	.505	.978	.014	1.44	.231	1.014	-.058	74.09	0.000	.944
(Scale)	1 ^a				1 ^a				1 ^a			

^a Fixed at the displayed value.

Even though the interpretation of the coefficients (θ) in the model may seem difficult due to the nature of the log link function $y = \exp(a + bx)$, a Poisson regression models the log of the expected g-index as a function of the predictor variables. The signs of the coefficients show whether the predictors have a positive or negative association with the g-index. A positive coefficient for a continuous variable indicates a positive relationship between the predictor and the g-index, while a negative coefficient indicates an inverse relationship. Our results suggest a lower association between the g-index and network-based measures in the field of Statistics, since only two variables show a statistically significant relationship with the g-index (as against three variables in Nanoscience and five in Pharmacology) and the goodness of the fit of the final model is weaker in this field.

Degree and tie strength are the variables which display the strongest relationship with the g-index in all three fields under study. Our results suggest that the degree is the most influential variable in Statistics (32% increase in g-index for every unit increase in

degree) while both degree and tie strength show similar influence in Nanoscience (around 22% increase in g-index) and tie strength is the most influential variable in Pharmacology (29% increase in g-index).

Local cohesion as measured by the clustering coefficient shows a negative association with the g-index both in Nanoscience and Pharmacology. This suggests that establishing collaborations with scientists who do not collaborate between them (for example, if they belong to different teams or work in different research lines within a team) is on average positive for the research performance of a given author. On the other hand, considering authors with the highest propensity to form cliques (clustering coefficient = 1), those in Statistics were more likely to obtain a high degree and a high g-index than those in the two experimental fields, a result that would suggest the lower negative effect of close networks on research performance in the theoretical field.

Closeness and eigenvector variables are significant only in Pharmacology. The positive association between closeness and the g-index suggests that global centrality is on average positive for research performance in the field, probably increasing the opportunity to find new collaborators. From a detailed examination of our data we observed that in Pharmacology the highest values of closeness centrality correspond to several scientists affiliated to hospitals who collaborate within their institution but also with scientists in other hospitals as well as in universities and pharmaceutical companies. This position, close to many other actors in the network, allows them to participate in highly relevant pharmacological research, such as that conducted in the framework of clinical trials and, in the long term, to obtain high g-index values.

The negative relationship between the eigenvector and the g-index observed in Pharmacology is somewhat counterintuitive, since better performance could be expected for the best connected authors. In fact, a positive bivariate correlation between the g-index and the eigenvector was observed (Spearman's $\rho = 0.218$). To explain this, we should keep in mind that the beta values in the multiple regression need to be understood in the context of the overall model. Other variables in the model may account for part of the information provided by the eigenvector in such a way that the negative beta value of the eigenvector corrects for an excessive positive influence of another related variable/s. In fact, the negative sign of the eigenvector implies a higher reduction of the g-index for those authors with the highest eigenvector values. This is consistent with our data, since the authors with the highest eigenvector values do not present the highest g-index values. The reason is that a high degree is very relevant to obtain a high g-index and for authors with a very high number of collaborators, it can be difficult to obtain a high eigenvector value, since it is unlikely for all the co-authors of a given author to be well-connected ones.

To gain further insight into the comparative importance of the number of collaborators and the strength of ties on the research performance of scientists, four categories of scientists were distinguished according to their values of degree (high or low) and strength of ties (high or low)²⁶. We observe that the g-index differs based on the four-group classification of scientists (Kruskal–Wallis test < 0.001). Figure 8.1 shows that the g-index tends to increase from authors with low degree-low strength (group 1) to those with high degree-high strength (group 4) in the three fields. Cumulative positive effects of degree and strength of ties can be observed in the fourth category. However, only in Nanoscience the differences between the g-index of the two intermediate categories are statistically significant ($p < 0.001$): g-index values tend to be higher for authors with high degree and low strength compared to those with low degree and high strength, thus suggesting that a diversity of links may outweigh the negative effect of low strength in that field.

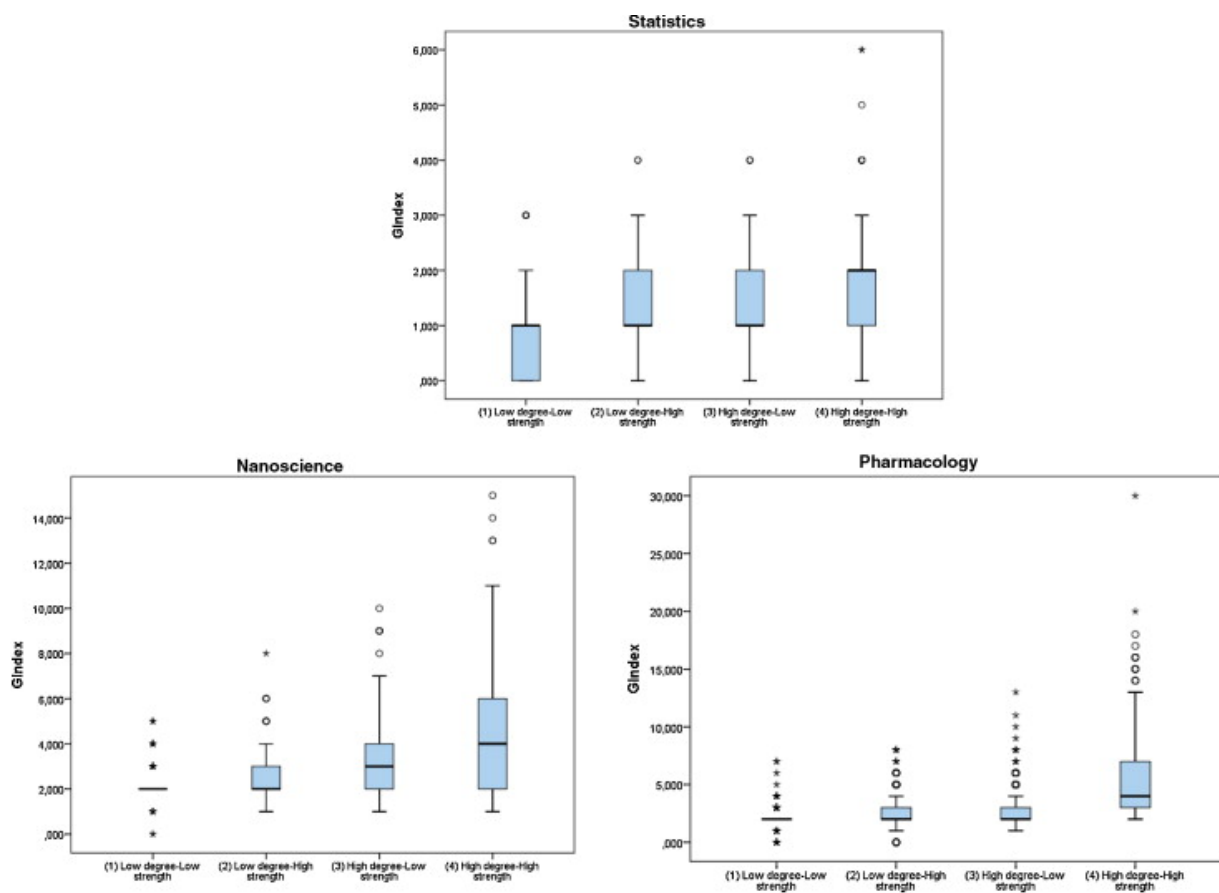


Figure 8.1. G-index by level of degree and strength of ties for authors in the three fields.

Note: significant differences between all pair of groups in Nanoscience ($p < 0.01$); between all pairs of groups ($p < 0.001$) except 2 vs 3 in Pharmacology; between groups 1 vs.3 and 1 vs.4 in Statistics ($p < 0.01$).

²⁶ The median is considered in each field to separate “low” values (\leq median) from “high” values ($>$ median).

8.5. Discussion

This study shows differences in the structure of the collaboration networks in the three fields under analysis. At the macro level, Pharmacology and Nanoscience present a similar network structure, denser than Statistics, which displays a less connected and more fragmented network. The underlying reason is the bigger size of teams in laboratory-based research conducted in experimental fields such as Pharmacology and Nanoscience when compared to Statistics, where scientists are more likely to work alone or in small teams.

The study at the micro-level confirms that there is a relationship between the position of Spanish scientists in co-authorship networks and his/her research performance as measured by the g-index. This association varies by field and seems to be stronger in Pharmacology and Nanoscience than in Statistics.

In all three fields under study, the variables which show a stronger relationship with the g-index are the average strength and the standardised degree. Specifically, scientists who have many collaborations with different scholars (high degree) or that build strong links with their co-authors (high average strength) are more likely to show a higher g-index. Among the benefits of a higher number of direct ties (high degree), knowledge sharing through interaction and discussion can be mentioned. Scientists can learn from one another and produce better research if they pool their knowledge, skills and resources (see for example, Abbasi et al., 2011 and Badar, Hite, & Badir, 2013). On the other hand, repeated co-authorships may be accounted for by mutual confidence and a set of shared norms of behaviour between the partners, which can facilitate resource sharing and cooperation (see for example, Abbasi et al., 2011, Ahuja, 2000 and Guler & Nerkar, 2012). In our study, there is a positive association between the g-index and both the number of collaborators and the strength of links, but the diversity of co-authors seems to be more important in Statistics, while the strength of the relationships with existing co-authors seems to be more relevant in Pharmacology. Strong ties can be more relevant in experimental fields such as Pharmacology due to the closer interaction and reciprocal support among members needed to conduct laboratory work. This is probably not so evident in Nanoscience – in spite of its experimental nature – because of the higher importance of diversity of sources (degree) in the more interdisciplinary fields (Jansen et al., 2010). In other fields, such as Information Systems and Information Technology, expanding social relationships, especially with different co-authors (Abbasi et al., 2011) but also with the same co-authors (Liao, 2011), also emerged as an effective way to improve research performance. Interestingly, authors with a high number of collaborators and strong ties show in our study the highest g-index values across all fields, although this was not the most common situation. Comparing the g-index of authors with a high

number of collaborators but low tie strength values with those showing a low number of collaborators but high tie strength values, significant differences were only found in Nanoscience, confirming the higher benefit drawn by the first set of authors. This finding is consistent with the important role played by the diversity of links in the more interdisciplinary fields above mentioned.

As regards the ongoing debate about which type of network (closed or open) is more beneficial for performance, we were not able to approach the subject through the study of constraint, a common feature in the literature, since this variable was removed from the analysis due to multicollinearity problems. However, our data point to a negative association between g-index and local cohesion (clustering coefficient) in the experimental fields, which means that widening the network of collaborators to scientists who do not collaborate between them is on average positive, at least in Pharmacology and Nanoscience, thereby suggesting that more open structures would be more beneficial in these fields.

Being a well-connected author (as measured by the eigenvector) is not associated to a higher g-index in two fields, while a negative association is observed in Pharmacology. An inverse relationship between research performance and the eigenvector was also reported by Abbasi et al. (2011) in a study on social networks in Information Science. The fact that well-performing scientists (i.e. research leaders of teams) had a great proportion of their papers written in collaboration with students rather than with other well-performing scientists was the explanatory reason held for this. In our study, we have observed that it can be especially difficult for scientists with a very high g-index to obtain a high eigenvector value, since they usually have many collaborators and it is unlikely for all their co-authors to be well-connected. The essential role of teams in Pharmacology, which include members in different stages of their scientific career (from students to senior scientists) and with different levels of productivity, needs to be considered to understand that collaboration limited to well-connected authors is not the norm in the field even for senior scientists, who maintain links with authors which may differ largely in their structural positions within the network.

The benefits of being, geodesically speaking, between many authors (high betweenness) has been reported in the literature (Li et al., 2013 and Yan & Ding, 2009), and are mainly based on the fact that these scientists have ties connecting otherwise disconnected authors thus enabling access to diverse sources of knowledge. However, our data suggest that in the case of Spanish scientists, playing a bridging role is not associated to a higher g-index in any of the three fields under analysis. As stated by Abbasi et al. (2011), “brokerage positions” may have strategic value, but do not necessarily improve research performance, maybe due to the costs of maintaining collaboration with authors from different contexts.

In summary, our study shows there is a relationship between the position of scientists in the co-authorship network and their research performance, with these relationships being stronger in the experimental fields (Pharmacology and Nanoscience) than in Statistics. Having a high number of collaborators and/or high strength of links with co-authors is associated to a higher g-index of scientists in all three fields. Including collaborators from different contexts, who do not collaborate between them, is also found to be a positive factor in all fields with the exception of Statistics. Being close to all other authors in the network is significant in Pharmacology, because these central positions are occupied by scientists who connect teams from different institutions and participate in highly relevant and collaborative research. No clear benefits from intermediary positions (high betweenness) or from those in connection with well-connected authors (high eigenvector centrality) are derived from this study.

Our research is subject to a series of limitations. (1) Firstly, our study is based on co-authorship, which is a partial indicator of scientific collaboration (Katz & Martin, 1997), so not all collaborative links between scientists are considered (e.g. those collaborators mentioned in the acknowledgment section or not mentioned at all are not visible). (2) We have used a single measure to assess research performance which indeed is a multidimensional endeavour and would require more complex measures. Moreover, the specific limitations described in the literature for the g-index, such as being affected by an occasional “big hit” (a highly cited document) (Costas & Bordons, 2008), apply to our study as well. (3) Conclusions derived from the association between the g-index of the authors and their network-based measures should be read with caution, since regression analyses reveal relationship among variables but do not imply that the relationships are causal, and the influence of other unmeasured variables cannot be discarded. (4) We do not know to what extent the delimitation of the fields according to WoS subject categories could influence the results, for example underestimating the interdisciplinarity of the fields. (5) Our results describe the behaviour of Spanish authors in three different fields, and may not be extrapolated to other communities of scientists.

Acknowledgments

This research was supported by the Spanish Ministry of Science and Innovation (MICINN) (research project CSO2008-06310) and the Spanish National Research Council (JAE predoctoral grant and project 201110E087). We are very grateful to Laura Barrios and José Manuel Rojo for their statistical advice as well as to two anonymous referees for their valuable comments.

References

- Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: a correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, 5, 594-607.
- Abbasi, A., Chung, K.S.K., & Hossain, L. (2012). Egocentric analysis of coauthorship network structure, position and performance. *Information Processing and Management*, 48(4), 671-679.
- Abramo, G., D'Angelo, C.A., & Di Costa, F. (2009). Research collaboration and productivity: is there correlation? *Higher Education*, 57, 155-171.
- Ahuja, G. (2000). Collaboration networks, structural holes and innovation: A longitudinal study. *Administrative Science Quarterly*, 45(3), 425-455.
- Badar, K., Hite, J.M., Badir, & Y.F. (2013). Examining the relationship of coauthorship network centrality and gender on academic research performance: the case of chemistry researchers in Pakistan. *Scientometrics*, 94, 755-775.
- Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica a-Statistical Mechanics and Its Applications*, 311(3-4), 590-614.
- Batagelj, V., Mrvar, A. (2013). Pajek. V 3.14. <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>.
- Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification *Journal of Mathematical Sociology*, 2(1), 113-120.
- Bordons, M., Aparicio, J., & Costas, R. (2013). Heterogeneity of collaboration and its relationship with research impact in a biomedical field. *Scientometrics*, 96 (2): 443-466.
- Bordons, M., & Gómez, I. (2000). Collaboration networks in science. In: B.Cronin and H.B.Atkins (Eds.) *The web of knowledge: A festschrift in honor of Eugene Garfield* (pp.197-213). Medford, NJ: Information Today.
- Borgatti, S.P., Mehra, A., Brass, D.J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323, 892-895.
- Burt, R.S. (1992). *Structural holes*. Cambridge, MA: Harvard University Press.
- Burt, R.S. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110(2), 349-399.
- Coleman, J.S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94(Supplement), 95-120.
- Costas, R., & Bordons, M. (2007). Algorithms to solve the lack of normalization in author names in bibliometric studies. *Investigación bibliotecológica*, 21(42), 13-32.
- Costas, R., & Bordons, M. (2008). Is g-index better than h-index? An exploratory study at the individual level. *Scientometrics*, 77(2), 267-288.
- De Nooy, W., Mrvar, A., & Batagelj, V (2005). *Exploratory network analysis with Pajek*. Cambridge: Cambridge University Press.

- Egghe, L. (1991). Theory of collaboration and collaborative measures. *Information Processing and Management*, 27, 177-202.
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152.
- Freeman, L.C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, 1(3), 215-239.
- Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1), 69-115.
- Glänzel, W., Schubert, A. (2004). Analysing scientific networks trough co-authorship. In: Moed, H.F., Glänzel, W. and Schmoch, U. Ed. Handbook of Quantitative S&T Research. Dordrecht: Kluwer Academic Publisher, pp. 257-276.
- Guler, I., & Nerkar, A. (2012). The impact of global and local cohesion on innovation in the pharmaceutical industry. *Strategic Management Journal*, 33, 535-549.
- Hanneman, R.A., & Riddle, M. (2005). Introduction to social network methods. University of California at Riverside: Riverside, CA.
- He, B., Ding, Y., & Ni, C. (2011). Mining enrich contextual information of scientific collaboration: a meso perspective. *Journal of the American Society for Information Science and Technology*, 62(5), 831-845.
- Heinze, T., & Bauer, G. (2007). Characterizig creative scientists in nano-S&T: Productivity, multidisciplinary, and network brokerage in a longitudinal perspective. *Scientometrics*, 70(3), 81-830.
- Hirsch, J.E. (2005). An index to quantify and individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572.
- Hou, H., Kretschmer, H., & Liu, Z. (2008). The structure of scientific collaboration networks in Scientometrics. *Scientometrics*, 75(2), 189-202.
- Jansen, D., Von Görtz, R., & Heidler, R. (2010). Knowledge production and the structure of collaboration networks in two scientific fields. *Scientometrics*, 83(1), 219-241.
- Katz, S., & Martin, B.R. (1997). What is research collaboration? *Research Policy*, 26, 1-18.
- Kleinberg, J.M. (1999). Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46(5), 604-632.
- Klenk, N.L., Hickey, G.M., & MacLellan, J.I. (2010). Evaluating the social capital accrued in large research networks: the case of the Sustainable Forest Management Network (1995-2009). *Social Studies of Science*, 40(6), 931-960.
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation* 11(1): 3-15.
- Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, 35, 673-702.
- Li, E.Y., Liao, C.H., & Yen, H.R. (2013). Co-authorship networks and research impact: a social capital perspective. *Research Policy*, 42, 1515-1530.

- Liao, C.H. (2011). How to improve research quality? Examining the impacts of collaboration intensity and member diversity in collaboration networks. *Scientometrics*, 86, 747-761.
- Li-Chun, Y., Kretschmer, H., Hanneman, R.A., & Ze-Yuan, L. (2006). Connection and stratification in research collaboration: an analysis of the COLLNET network. *Information Processing and Management*, 42, 1599-1613.
- Moed, H. F. (2005). Citation analysis in research evaluation. Dordrecht: Springer.
- Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital and the organizational advantage. *Academy of Management Review*, 23(2), 242-266.
- Newman, M.E.J. (2001). The structure of scientific collaboration networks. *PNAS*, 98(2), 404-409.
- Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*, 28, 443-455.
- Reagans, R., & Zuckerman, E.W. (2001). Networks, diversity and productivity: the social capital of corporate R&D teams. *Organization Science*, 12(4), 502-517.
- Scott, J. (1991). Social network analysis. A handbook. London: Sage Publications.
- Sonnenwald, D.H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, 41(1), 643-681.
- Vinkler, P. (2010). The evaluation of research by scientometric indicators. Oxford: Chandos Publishing.
- Walker, G., Kogut, B., & Shan, W. (1997). Social capital, structural holes and the formation of an industry network. *Organization Science*, 8, 109-125.
- Wasserman, S., & Faust, K, (1994). Social network analysis: Methods and applications. Cambridge: Cambridge University Press.
- Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: a co-authorship network analysis. *Journal of the American Society for Information Science and Technology*, 60(10), 2107-2118.