Information Processing and Management 48 (2012) 671–679

Contents lists available at SciVerse ScienceDirect

Information Processing and Management

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

Egocentric analysis of co-authorship network structure, position and performance

Alireza Abbasi *, Kon Shing Kenneth Chung, Liaquat Hossain

Centre for Complex Systems Research, Faculty of Engineering and IT, University of Sydney, NSW 2006, Australia

ARTICLE INFO

Article history:
Received 27 May 2011
Received in revised form 6 September 2011
Accepted 16 September 2011
Available online 19 October 2011

Keywords:
Co-authorship network
Social network analysis
Ego-centric analysis
Scholars performance
g-Index

ABSTRACT

In this study, we propose and validate social networks based theoretical model for exploring scholars' collaboration (co-authorship) network properties associated with their citation-based research performance (i.e., g-index). Using structural holes theory, we focus on how a scholar's egocentric network properties of density, efficiency and constraint within the network associate with their scholarly performance. For our analysis, we use publication data of high impact factor journals in the field of "*Information Science & Library Science*" between 2000 and 2009, extracted from Scopus. The resulting database contained 4837 publications reflecting the contributions of 8069 authors. Results from our data analysis suggest that research performance of scholars' is significantly correlated with scholars' ego-network measures. In particular, scholars with more co-authors and those who exhibit higher levels of betweenness centrality (i.e., the extent to which a co-author is between another pair of co-authors) perform better in terms of research (i.e., higher g-index). Furthermore, scholars with efficient collaboration networks who maintain a strong co-authorship relationship with one primary co-author within a group of linked co-authors (i.e., co-authors that have joint publications) perform better than those researchers with many relationships to the same group of linked co-authors.

1. Introduction

As in most large organizations, performance of individuals and teams is measured through a set of metrics that pertain to task and contextual performance. Similarly in academia, scholars and scientists are evaluated based on their academic performance such as teaching evaluations, governance capabilities, research output, number of secured grants and so on. Such evaluation of researchers is not only needed for faculty recruitment, but also for governmental funding allocation and for achieving a high reputation within the research community. The reputation of research organizations indirectly affects the society's welfare, since a high reputation attracts foreign purchases, foreign investments, and highly qualified students from around the world. The implication of such ranking provides basis and justification for federal funding thus encouraging high research standards and goals. Therefore, on a global level, with respect to governmental funding (i.e., the allocation of funding for a specific project to a scientific research group) and university strategy, it is important to identify key scholars, collaboration areas and research strengths within universities with the aim of maximizing the research output, cost optimization, and resource utilization. However, in all these cases, the common problem exists, namely answering the question of how can research productive scientists be identified, clustered, and configured for optimal research synergies (Jiang, 2008).

* Corresponding author.
E-mail address: alireza.abbasi@sydney.edu.au (A. Abbasi).

0306-4573/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ipm.2011.09.001

In light of the above, in recent years there has been a sharp increase in the number of collaborations between scholars. An explanation for the rapid growth of international scientific collaboration has been provided by Luukkonen, Persson, and Sivertsen (1992) and Luukkonen, Tijssen, Persson, and Sivertsen (1993) as well as Wagner and Leydesdorff (2005). By jointly publishing a paper, researchers show their knowledge sharing activities, which are essential for knowledge creation. As most scientific output is a result of group collaboration, it often needs scientific cooperation between individuals across national borders (Leclerc & Gagné, 1994).

Due to the necessity to keep pace with scientific progress not only at the micro level (e.g., level of individual researchers) but also at the macro level (i.e., nationally), most governments are interested in enhancing the level of international collaborations through appropriate policies (Katz & Martin, 1997; van Raan, 2004). Scientific collaboration in addition to advance research, facilitates increasing the visibility and authorship of the highly productive researchers (Pao, 1992). An important result of scientific collaborations is the creation of new scientific knowledge, including new research questions, new research proposals, new theories, and new publications (Stokols, Harvey, Gress, Fuqua, & Phillips, 2005). Although Duque et al. (2005) have found that collaboration was not associated with an increase in scientific publications in the developing countries of Ghana, Kenya, and India (Kerala) (2005), Lee and Bozeman show that the total number of publications for US scientists is positively associated with the total number of collaborations (Lee & Bozeman, 2005). Also, other researchers show that research collaboration enhance the quality of research (considering publications' citation count) in different disciplines such as medicine (Figg et al., 2006); biotechnology and applied microbiology (Frenken, Holzl, & Vor, 2005); and chemistry (Glänzel & Schubert, 2001).

Since scientific collaborations are defined as "interactions taking place within a social context among two or more scientists that facilitates the sharing of meaning and completion of tasks with respect to a mutually shared, super-ordinated goal" (Sonnenwald, 2007), those collaborations frequently emerge from, and are perpetuated through social networks – formal and informal. Since social networks may span disciplinary, organizational, and national boundaries, social networks can influence collaboration in multiple ways (Barabasi et al., 2002; Cross, Borgatti, & Parker, 2002; Kraut, Egido, & Galegher, 1988; Newman, 2004; Sonnenwald, 2007). Co-authorship network represents a prototype of a social network by mapping the graph containing authors who have joint relevant publications (Yin, Kretschmer, Hanneman, & Liu, 2006).

To date, most studies on the effects of collaboration network properties and its evolution on citation patterns have either focused on (i) sociocentric properties of centralization, cohesion (e.g., clique structure) and density or (ii) egocentric properties of node centrality such as betweenness, point and closeness centrality. Very few though, have considered evaluating the position effect of egocentric network constraint and efficiency on performance in the scholarly collaboration network domain. The motivating questions for our study are: (i) what predictors of egocentric network position explain scholarly performance within a collaboration network? (ii) can structural holes theory, its assumptions and its implication for individual performance be applied and extended within the context of a non-competitive, non-corporate and educational settings? (iii) do network position properties matter more than properties of network structure in terms of scholarly performance? (iv) what social factors and implications should be accounted for scholars in order to enhance performance in an educational setting?

For our analysis, we use publication data of high impact factor journals in the field of "*Information Science & Library Science*", extracted from Scopus. In this study, we analyse the results for co-authorship networks of the authors who has at least one publication in the top 9 selected journals of the field. For the purpose of this study, single-authored papers were neglected as part of the analysis.

In the following section, we review current literature on theories of social network and performance. In particular, we focus on egocentric network literature such as point centrality and structural holes theory and present a model to understand individual scholarly performance. Section 3 describes the data resources, the data gathering and validation process followed by a definition of measures used in our analysis. Section 4 provides details of our findings. Correlation results of network properties and individual performance are also presented. Finally, we discuss the implication of the results, research limitations, and our future work.

2. Literature review

2.1. Overview of social network

A social network is a constituent of two or more actors (individuals) who are connected through one or more relationships such as providing advice, information and so on. In general, analysis of social networks is usually either conducted using the sociocentric approach or the egocentric approach (Chung, Hossain, & Davis, 2005). In the latter, the node of interest is the ego, and its immediate neighbors are the alters. At the individual level, the debate concentrates on how the structural position of an individual in the network impacts outcome, such as performance, of that person (Chung & Hossain, 2009).

Social networks operate on many levels, from families up to the level of nations. They play a critical role in determining the way problems are solved, organizations are run, markets evolve, and the degree to which individuals succeed in achieving their goals (Abbasi & Altmann, 2011). Social networks have been analyzed to identify areas of strengths and weaknesses within and among research organizations, businesses, and nations as well as to direct scientific development and funding policies (Owen-Smith, Riccaboni, Pammolli, & Powell, 2002; Sonnenwald, 2007).

In general, the benefit of analyzing social networks is that it can help people to understand how to share professional knowledge in an efficient way and to evaluate the performance of individuals, groups, or the entire social network. For instance, with respect to performance evaluation, the social network of a researcher within a research community can be considered an indication of his collaboration activity (Abbasi, Altmann, & Hwang, 2010).

2.1.1. Networks structure and performance

One of the earliest studies that related sociometric aspects of human communication patterns to performance was that of the 'Bavelas–Leavitt Experiment' (Bavelas, 1950; Leavitt, 1949), also known as the MIT experiment. Inevitably, a thought-provoking finding that emerged from this study back then was that centralized structures (e.g., star (or wheel) network) were far more conducive to performance (solving the puzzle faster) in contrast to decentralized or flatter structures (e.g., circle network). The crux of the argument is that information flow is inefficient in decentralized networks and therefore less conducive to performance. However, later research by Guetzkow and Simon (1955) revealed that decentralized structures actually worked better than centralized structures when tasks become more complex.

2.1.1.1. Freeman's concept on centrality. As described, the idea of centrality was applied to human communication in the early 1950s. Although the 'structural centrality and influence in group processes' hypothesis was proposed by Bavelas (1950) and reported in depth by Leavitt (1951), modifications and extension to the original experiments provided contradictory and confusing results (Burgess, 1968). In the late 1970s, Freeman (1978) wrote a seminal article that clarified the conceptual foundations of centrality, which soon became a core concept in social network studies. His work laid the foundation for social network scholars to apply and extend the notion of structural centrality at both the node and network level, conceptually and empirically.

Freeman (1978) reviewed the various measures and overlapping concepts of centrality by unifying the measures while clarifying its range and limits of potential for application. In particular, he defined centrality in terms of point, betweenness and closeness centrality, with each having important implications on social outcome and processes. According to Freeman (1978), "these kinds of centrality imply three competing 'theories' of how centrality might affect group processes". Degree centrality can be measured in terms of degree (the number of ties to and from an actor) and looking at local (ego) structure of an actor, centrality is measured in terms of ego-betweenness (the extent to which an actor lies in the shortest path to all alters). Each centrality concept has been related to important social occurrences – degree centrality being viewed as an important indicator of an actor's communication activity and ego-betweenness centrality as an indicator of the potential of an actor's control of communication.

In a subsequent study, Freeman, Roeder, and Mulholland (1979) reverted to the classic experiment by Bavelas (1950) to study the effects of structural centrality on human communication. Using 100 volunteers (university students) as subjects for the experiment, Freeman et al. (1979) analyzed the results and demonstrated that centrality is an important structural factor influencing leadership, satisfaction and efficiency. In particular, out of the all concepts of structural centrality, only two demonstrated interesting results and significance in its effect on performance; namely the control based measure of betweenness and the activity based measure of degree. Interestingly, another structural factor, the overall density of communication paths in the structural form also turned out to be relevant in understanding performance.

In another classical work on the effects of network structure on innovation diffusion, Coleman, Katz, and Menzel (1957) attempted to understand the underlying social processes that affected 125 doctors' rate of adoption of a new drug. Results suggested that doctors who were generally more integrated with their peers, that is, with denser networks, were faster in the adoption of the new drug as compared to those who were more isolated. The results from Coleman et al.'s (1957) study suggests that the larger number of ties an individual has (higher density) results in a higher likelihood to diffuse innovation faster. These results resonated strongly with similar findings about the density concept by Freeman et al. (1979) described above.

In line with the arguments above, it is expected that authors, which have a high degree centrality and a high betweenness centrality, may have a high potential of a good research performance as they are in the center of network. As detailed below, we can formulate these expectations as the following three hypotheses:

H1. A researcher's collaboration network structure correlates to performance.

H1a. Degree centrality of a researcher's collaboration network positively correlates to performance.

H1b. Ego-betweenness centrality of a researcher's collaboration network positively correlates to performance.

2.1.1.2. Burt's structural holes theory. In the early 1990s, Burt (1992) made an influential contribution to the network paradigm and phenomena of structural effects on individual outcome by shifting the focus from network structure and network relations to network position. Burt's (1992) theory on structural holes offers a novel and subtle but interesting perspective in explaining why some individuals perform better and others do not. For example, it takes Coleman et al.'s (1957) study and its assumptions a step further by offering an explanation of why social processes such as innovation diffusion may occur faster from a structural positional point of view rather than from a relational perspective. The theory is linked to personality theory suggesting that personal attributes (such as locus of control, leadership skills, ability to perform well) of an individual is associated with structural autonomy – an optimal situation where an individual benefits from non-redundant information benefits.

In contrast to network structure and relational ties, Burt (1992) argues that structural configuration of an individual's social network, which provides optimized "brokerage" position is what dictates structural advantages such as information novelty and control. The basis for this argument leverages on the fact that maximizing the number of ties (weak or strong) in an individual's network does not necessarily provides such benefits. Instead, opportunity costs come into play and maintenance of ties become expensive in terms of time and resource. Furthermore, as an individual's personal network grows over time, the extent of information coming from closely knit clusters tend to become redundant. To this end, Burt (1992) capitalizes on his theory of structural holes by focusing on the importance of structural position rather than structural properties such as density or the size of the network.

'Holes' in the network is the absence of ties, which would otherwise connect unconnected clusters together. Individuals who bridge these holes attain an advantageous position that yields information and control benefits. Therefore, structural holes theory is based on the idea that actors are in a better position to benefit from interactions with others if actors are connected to others who are not well-connected themselves or well-organized. In other words, the bridging of connection to others provides opportunities; the lack of connections among those others is the holes in the structure (and therefore, structural holes).

Closer examination of structural holes theory reveals that it is based on the assumption of betweenness centrality: that power and influence accrue to those who broker connections between unconnected groups of people. It is this concept of betweenness centrality (or ego-betweeness in the case of egocentric networks) that Burt (1992) capitalizes on and extends to explain the role of "brokerage" as a form of obtaining structural autonomy which leads to improved performance, getting ahead and obtaining good ideas. This theoretical contribution offers a more insightful perspective on individual performance given that Guetzkow and Simon (1955) note that centrality in itself is not always a key predictor of individual performance. Instead, the theory offers insightful explanation beyond the concept of centrality and centralization in that an individual's benefit accrues from the extent that individual's network is efficient, effective and constrained. Therefore, we hypothesize that:

H1c. Density of a researcher's collaboration network negatively correlates to performance.

2.1.1.3. Network efficiency and effectiveness. In order to optimize a network by capitalizing on structural holes, Burt (1992) claims that increasing network size (number of direct contacts) without considering the diversity reached by the contacts makes the network inefficient in many ways. Therefore, the number of non-redundant contacts is important to the extent that redundant contacts would lead to the same people and hence provide the same information benefits. The term "effectiveness" is used to denote the average number of people reached per primary contact; while the term "efficiency" concerns the total number of people of people reached with all primary contacts. Hence, effectiveness is about the yield per primary contact while efficiency is about the yield of the entire network. To illustrate, the network diagrams in Fig. 1 contrasts an inefficient network (A) to an efficient network (B).

The term that Burt (1992) uses to denote effectiveness in networks is effective size. In network A, the network size is 16. The effective size however, is 4, because in effect, the ego is only able to obtain novel information and benefits from the four clusters, which is not connected to each other except through 'you'. The other three ties to each of the cluster are redundant because they provide the same information that is available through the fourth. Efficiency in network A is therefore 0.25 (measured as effective size (4)/network size (16)). In network B, the network size is 4, and effective size is 4, resulting in perfect efficiency of 1 (4/4). Ideally, the number of non-redundant contacts should increase with the number of contacts to achieve optimal efficiency (i.e., 1). As one increases one's number of contacts and gradually start to have a smaller number of non-redundant contacts, the individual's network efficiency decreases. Conversely, as the number of non-redundant contacts increases relative to the lower number of contacts, the individual's network efficiency increases.

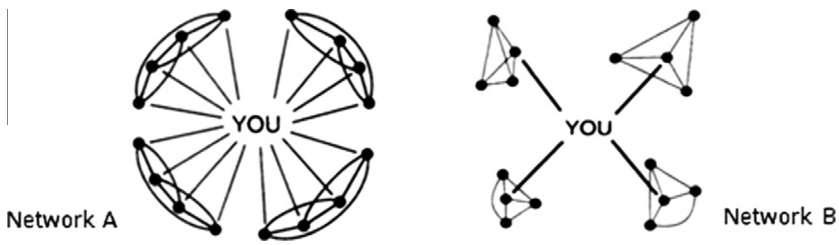

Fig. 1. Inefficient (A) and efficient (B) networks (adapted from Burt (1992)).

2.1.1.4. Network constraint. Constraint dictates the extent to which an individual's opportunities are limited by investing the bulk of his or her network time and energy in relationships that lead back to the single contact (Burt, 1992). In other words, constraint measures the degree to which an individual's contacts are connected to each other and is therefore a proxy for redundancy of contacts. According to Hanneman (2001), constraint also measures the extent to which an ego is connected to others who are connected to one another. So if the ego has many connections to others who in turn has many connections to more others, the ego is quite constrained. At organizational levels, individuals with high constraint indices are unable to conceive novel ideas because of the redundant nature of information that is sourced from a densely connected group of individuals. Previous research has consistently demonstrated that high efficiency and low constraint indices are useful indicators of an individual's ability to produce good ideas (Burt, 2004), "getting ahead" in terms of job performance and promotion (Burt, 1992, 2005) and enjoy greater career mobility (Podolny & Baron, 1997).

With respect to our study, a disjunct group of primary contacts relates to co-authors that have joined publications (i.e., that are linked). Therefore, testing this property means testing whether a scholar maintains strong relationships with all co-authors of a group of linked co-authors or whether the scholar focus on a strong relationship with just one co-author of this group. On the other hand, it is expected that scholars who have more redundant contacts (which means higher constraints) will received redundant information that leads to weaker performance. Therefore, in order to test this property, we hypothesis the following:

H2. Network position of a researcher's collaboration network correlates to performance.

H2a. Efficiency of a researcher's collaboration network positively correlates to performance.

H2b. Constraints of a researcher's collaboration network negatively correlates to performance.

3. Data and method

3.1. Data collection

For our analysis, we selected publication data of top 9 journals in "Information Science & Library Science" based on their 2009 impact factor as enumerated in the current edition of *Journal Citations Report*.[1] To construct our database, we extracted publication information of the top 9 journals (i.e., *MIS Quarterly, Journal of American Medical Informatics Association, Journal of Computer-Mediated Communication, Journal of Informetrics, Annual Review of Information Science & Technology, International Journal of Computer-Supported Collaborative Learning, Journal of American Society for Information Science & Technology, Information & Management, and Scientometrics*) from Scopus, as one of the main sources which present bibliometric data. We restricted our query to just articles and reviews and publish year between 2000 and 2009. We ignored the output which their authors and affiliation records were empty. While the list includes top 10 journals but unfortunately, we could not get the data of the "*Journal of the Association for Information Systems*" (Rank 9).

After extracting the publications meta-data, we stored publication information (i.e., title, publication date, author names, affiliations, publisher, number of citations, etc.) in a relational database and extract relationships (e.g., co-authorships) between researchers and stores the data in the format of tables in its local database. Different types of information were extracted from each publication meta-data: Publications information (i.e., title, publication date, journal name, etc.); Authors' names and Affiliations of authors (including country, institution and department names). The resulting database contained 4837 publications that received totally 93,660 citations, reflecting the contributions of 8069 authors and 12,783 collaborations.

3.2. Measures

3.2.1. Scholars' performance

To assess the performance of scholars, many studies suggest quantifying scholars' publication activities as a good measure for the performance of scholars. The general idea is that a researcher gets a high visibility in the research community, if the researcher publishes and her publications get cited. The number of citations qualifies the quantity of publications (Lehmann, Jackson, & Lautrup, 2006). Hirsch introduced the h-index as a simple measure that combines in a simple way the quantity of publications and the quality of publications (i.e., number of citations) (Hirsch, 2005). A scholar with an index of h has published h papers, which have been cited by others at least h times (Hirsch, 2005).

Although other kinds of indicators for scholars' performance (impact) has been proposed such as *AuthorRank* by (Liu, Bollen, Nelson, & Van de Sompel, 2005) and weighted PageRank by (Yan & Ding, 2011), but the h-index became also the basis for a wide range of new measures (Altmann, Abbasi, & Hwang, 2009; Batista, Campiteli, & Kinouchi, 2006; Egghe, 2006; Jin, 2006; Sidiropoulos, Katsaros, & Manolopoulos, 2007; Tol, 2008). One of the most famous and widely used and accepted extension measure of h-index is g-index introduced by Egghe (2006) to overcome the main shortcomings of the h-index,

[1] http://sciencewatch.com/dr/sci/10/aug15-10_1

namely, ignoring the number of citations in excess of h. The g-index is defined as the average of citations of top g cited papers which have at least g citations each or in another words the largest number such that the top g cited papers receive together at least g^2 citations.

Although there is considerable debate on the reliability of the h-index and g-index (and its variants) (Haque & Ginsparg, 2009) the h-index and g-index are still widely used world-wide amongst academics. While the reliability of the measure is not the subject of this paper per se, it does provide at least an empirical metric so as to gauge a researcher's prolificacy. Thus, we will consider g-index as a citation-based surrogate measure as proxy for performance of research scholars.

3.2.2. Ego-network measures

3.2.2.1. Density. The number of ties in the ego network (not counting ties involving ego) divided by the number of pairs of alters in the ego network (i.e., potential ties).

3.2.2.2. Degree centrality. The number of actors (alters) that ego is directly connected to. It is also consider as size of ego network.

3.2.2.3. Ego-betweenness. The sum of ego's proportion of times ego lies on the shortest path between each part of alters. For alters connected to each other, the contribution to between of that pair is 0, for alters connected to each other only through ego, the contribution is 1, for alters connected through ego and one or more other alters, the contribution is $1/k$, where k is the number of nodes which connects that pair of alters.

3.2.2.4. Effective size. Burt's measure of the effective size of ego's network (the number of alters minus the average degree of alters within the ego network, not counting ties to ego network, not counting ties to ego).

3.2.2.5. Efficiency. The effective size divided by the number of alters in ego's network.

While the effective size of ego's network may tell us something about ego's total impact; efficiency tells us how much impact ego is getting for each unit invested in using ties. An actor can be effective without being efficient; and actor can be efficient without being effective.

3.2.2.6. Constraint. a measure of the extent to which ego is invested in people who are invested in other of ego's alters.

If two nodes (j and q) are alters (direct neighbors) of node i, the constraint (C_{ij}) of node i by node j is computed by the squares of the sum of the direct link strength and the indirect link strength from node i to node j (Burt, 1992):

$$C_{ij} = \left(p_{ij} + \sum_q p_{iq}p_{qj}\right)^2 (ji_{ne}, qi_{ne}, q \neq i, q \neq j) \tag{1}$$

3.3. Methods

We use social network analysis as a method to analyze collaboration network among scientists. Social networks are represented as a graph, which is constructed of nodes (actors or vertices) and links (ties, relations, or edges). Nodes, which denote individuals, organizations, or information, are linked, if one or more specific types of relationships (e.g., financial exchange, friendship, trade, and Web links) exist between them. For example, a node could represent a person, while a link between two nodes could represent that these two persons know each other in some way. Based on the co-authorships of publications of scholars, we construct the research collaboration network of scholars. Nodes of the research collaboration network represent scholars. A link between two nodes represents a publication co-authorship relationship between those scholars.

By calculating social network analysis (SNA) measures and one researcher productivity index (i.e., g-index), we aim to find whether the structure and position of a researcher within his local co-authorship network correlates with his research performance. Thus, after preparing the social network matrix, we used UCINET (Borgatti, Everett, & Freeman, 2002) as a tool for visualizing the network and for calculating network measures including degree centrality, ego-betweenness centrality, efficiency and constraints of each node. For correlating these measures and the research performance measure (i.e., g-index), we used the Spearman correlation values to measure association as our data was not normally distributed.

4. Analysis and results

Based on the available publication data of researchers, we built a network matrix for the co-authorship network. This matrix forms the basis for social network analysis. After importing the network matrix as a table into UCINET, we calculated the social network measures of our model and the citation-based performance measures (i.e., g-index) for all scholars. To illustrate, the results for the top 20 productive scholars are shown in Table 1.

Table 1
Top 10 productive scholars: Name, degree centrality, ego-betweenness centrality, density, efficiency, constraints, g-index and Cit_Sum.

	Name	Density	Degree	Ego-Betw	Efficiency	Constraints	g-Index	Cit_Sum
1	D.W. Bates	0.058	149	18380.9	0.948	0.036	38	2040
2	W. Glanzel	0.074	27	619	0.946	0.213	34	1269
3	M. Thelwall	0.06	30	795	0.979	0.089	31	1124
4	L. Leydesdorff	0.012	19	338	1.018	0.085	26	841
5	L. Egghe	0.111	10	74	0.869	0.493	25	696
6	R. Rousseau	0.043	36	1177	0.972	0.114	23	607
7	G. Hripcsak	0.122	51	1968.5	0.915	0.08	22	705
8	A. Schubert	0.132	14	149.7	0.866	0.436	22	553
9	H.-D. Daniel	0.311	10	48	0.73	0.538	21	475
10	J. Bar-Ilan	0.139	9	60	0.947	0.253	20	417

Table 2
Spearman correlation coefficients (rho) between scholars' performance measure and their ego-network measures ($N = 8069$).

	g-Index	Density	Degree	Ego-between	Efficiency
g-Index	–				
Density	−.538*	–			
Degree	.327*	.058*	–		
Ego-betweenness	.771*	−.646*	.465*	–	
Efficiency	.299*	−.076*	−.301*	.326*	–
Constraints	−.333*	.575*	−.484*	−.443*	.487*

* Correlation is significant at the 0.01 level (2-tailed).

Based on the data of Table 2 and the SNA results for all scholars, we test our hypotheses. In particular, we calculated the Spearman correlation coefficient between the ego-network measures explained and defined in Section 3.3 and the performance indices (g-index). The results are shown in Table 2.

As the results in Table 2 show, the Spearman correlation coefficient shows that there is very strong and significant positive correlation between scholars' network measure and performance. Therefore, we can reject each of the null hypotheses stipulated earlier.

Density shows a negative association with performance measures which means if a scholar's direct contacts are highly connected (denser ego-network) the performance of scholars will be lower. That is because highly connected neighbors form cluster and thus become redundant contacts for the scholar which lead to redundant information flow. On the other hand, the results indicate that scholars with more co-authors and those who exhibit higher levels of ego-betweenness centrality (i.e. the extent to which a co-author is between another pair of co-authors) perform better. Thus, we can infer that network structure of a scholar correlates to his/her performance (g-index).

The results related to efficiency and constraints show that scholars who maintain a strong co-authorship relationship to only one co-author of a group of linked co-authors perform better than those researchers with many relationships to the same group of linked co-authors. Therefore, we can infer that scholars' network position associates with his/her performance (g-index).

In brief, applying Spearman correlation rank test we found that, given that the results are statistically significant, there is sufficient evidence for us to reject the null hypothesis, therefore lending support to all our (alternative) hypotheses proposed. Therefore, we can infer that both *researchers' collaboration network structure* and *their position in the network* correlate to their citation-based performance (i.e., g-index).

5. Conclusion and discussion

As past research has shown, the h-index and g-index can be a surrogate measure for evaluating the research performance of scholars (Borgman & Furner, 2002). In addition to this, the collaboration skills of researchers became more and more important over the past years.

However, it still remains an open question as to whether the collaboration skills and research performance of researchers are correlated. In order to address this question, we used co-authorship data and social network analysis measures. The co-authorship data is used to derive the collaboration network of researchers. Ego-network measures of density, degree centrality, ego-betweenness centrality, efficiency and constraints were considered.

The results of our analysis show that research performance is positively associated with a researcher's ego-network structure (i.e., degree centrality, ego-betweenness centrality, density) and a researcher's position within the collaboration network (i.e. efficiency and constraint). Scholars connecting to more authors and the one who lies more on the shortest path between each of his co-authors demonstrate better research performance. With respect to efficiency, scholars who maintain strong co-authorship relationships to only one co-author of a group of linked co-authors (i.e., co-authors that have also joined publications) perform better than scholars with relationships to many co-authors within a group of linked co-authors. Thus, within the context of this research, we may suggest that scholars should rather than continuing collaboration with authors within the same cluster which lead to lower efficiency and higher constraint, ought to "bridge structural holes" by connecting with new and diverse research groups (e.g., inter-disciplinary groups) and fostering collaborations in order to improve research publication and prolificacy. In brief, results from our data analysis show that the research performance of scholars is significantly correlated with scholars' ego-network measures.

Furthermore, access to demographic information of researchers (e.g., age, gender, and nationality) would be useful as moderating variables in our proposed model. Based on this data, we would be able to categorize researchers and analyze the outcome for each of the demographic categories. This could help us to form a generalization of our model. The current lack of access to this kind of information can be considered a limitation of our research while at the same time an avenue for future research. Furthermore, using only one field of study (discipline) also can be considered as another limitation of our study, as it is probable that results vary based on different disciplines. Therefore, in order to have a generalized conclusion about the association between researchers' performance and their collaborative activity similar method to several disciplines should be applied in future studies.

References

Abbasi, A., Altmann, J., & Hwang, J. (2010). Evaluating scholars based on their academic collaboration activities: Two indices, the Rc-index and the Cc-index, for quantifying collaboration activities of researchers and scientific communities. *Scientometrics, 83*(1), 1–13.

Abbasi, A., & Altmann, J. (2011). On the correlation between research performance and social network analysis measures applied to research collaboration networks. In *Proceedings of the 44th annual Hawaii international conference on system sciences.* Waikoloa, HI: IEEE.

Altmann, J., Abbasi, A., & Hwang, J. (2009). Evaluating the productivity of researchers and their communities: The Rp-index and the Cp-index. *International Journal of Computer Science and Applications, 6*(2), 104–118.

Barabasi, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A, 311*(3–4), 590–614.

Batista, P., Campiteli, M., & Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics, 68*(1), 179–189.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *Journal of Acoustical Society of America* (22), 725–730.

Borgatti, S., Everett, M., & Freeman, L. (2002). Ucinet for windows: Software for social network analysis (Version 6), *Harvard, MA: Analytic Technologies*).

Borgman, C., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology* (36), 3–72.

Burgess, R. L. (1968). Communication networks and behavioral consequences. *Human Relations* (22), 137–159.

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

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology, 110*(2), 349–399.

Burt, R. S. (2005). *Brokerage and closure: An introduction to social capital*. New York: Oxford University Press.

Chung, K.S.K., Hossain, L., & Davis, J. (2005). Exploring sociocentric and egocentric approaches for social network analysis. In *International conference on knowledge management Asia Pacific*. Victoria University Wellington, New Zealand, pp. 1–8.

Chung, K. S. K., & Hossain, L. (2009). Measuring performance of knowledge-intensive workgroups through social networks. *Project Management Journal, 40*(2), 34–58.

Coleman, J. S., Katz, E., & Menzel, H. (1957). The diffusion of an innovation among physicians. *Sociometry, 20*(4), 253–270.

Cross, R., Borgatti, S. P., & Parker, A. (2002). Making invisible work visible: using social network analysis to support strategic collaboration. *California Management Review, 44*(2), 25–46.

Duque, R., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D., & Shrum, W. (2005). Collaboration paradox: Scientific productivity, the internet, and problems of research in developing areas. *Social Studies of Science, 35*(5), 755.

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics, 69*(1), 131–152.

Figg, W. D., Dunn, L., Liewehr, D. J., Steinberg, S. M., Thurman, P. W., Barrett, J. C., et al (2006). Scientific collaboration results in higher citation rates of published articles. *Pharmacotherapy, 26*(6), 759–767.

Freeman, L. C. (1978). Centrality in social networks: conceptual clarification. *Social Networks, 1*(3), 215–239.

Freeman, L. C., Roeder, D., & Mulholland, R. R. (1979). Centrality in social networks: Ii. Experimental results. *Social Networks, 2*(2), 119–141.

Frenken, K., Holzl, W., & Vor, F. (2005). The citation impact of research collaborations: The case of european biotechnology and applied microbiology (1988–2002). *Journal of Engineering and Technology Management, 22*(1–2), 9–30.

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

Guetzkow, H., & Simon, H. (1955). The impact of certain communication nets upon organization and performance in task-oriented groups. *Management Science, 1*(3/4), 233–250.

Hanneman, R.A. (2001). Introduction to social network methods. Retrieved 8th August, 2004 <http://faculty.ucr.edu/~hanneman/SOC157/NETTEXT.PDF>.

Haque, A., & Ginsparg, P. (2009). Positional effects on citation and readership in arxiv. *Journal of the American Society for Information Science and Technology, 60*(11), 2203–2218.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences, 102*(46), 16569.

Jiang, Y. (2008). Locating active actors in the scientific collaboration communities based on interaction topology analyses. *Scientometrics, 74*(3), 471–482.

Jin, B. (2006). H-index: An evaluation indicator proposed by scientist. *Science Focus, 1*(1), 8–9.

Katz, J., & Martin, B. (1997). What is research collaboration? *Research Policy, 26*(1), 1–18.

Kraut, R., Egido, C., & Galegher, J. (1988). Patterns of contact and communication in scientific research collaboration. In: *Proceedings of the 1988 ACM conference on computer-supported cooperative work*. Portland, Oregon, United States: ACM, pp. 1–12.

Leavitt, H. J. (1949). *Some effects of certain communication patterns on group performance*. Boston: Massachusetts Institute of Technology.

Leavitt, H. (1951). Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology* (46), 38–50.

Leclerc, M., & Gagné, J. (1994). International scientific cooperation: The continentalization of science. *Scientometrics, 31*(3), 261–292.

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

Lehmann, S., Jackson, A., & Lautrup, B. (2006). Measures for measures. *Nature, 444*(7122), 1003–1004.

Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing & Management, 41*(6), 1462–1480.

Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology & Human Values, 17*(1), 101.

Luukkonen, T., Tijssen, R., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics, 28*(1), 15–36.

Newman, M.E.J. (2004). Coauthorship networks and patterns of scientific collaboration. *PNAS* (101:suppl_1), April 6, 2004, pp. 5200–5205.

Owen-Smith, J., Riccaboni, M., Pammolli, F., & Powell, W. (2002). A comparison of US and European university-industry relations in the life sciences. *Management Science, 48*(1), 24–43.
Pao, M. L. (1992). Global and local collaborators: A study of scientific collaboration. *Information Processing & Management, 28*(1), 99–109.
Podolny, J., & Baron, J. (1997). Resources and relationships: Social networks and mobility in the work-place. *American Sociological Review, 62*(5), 673–693.
Sidiropoulos, A., Katsaros, D., & Manolopoulos, Y. (2007). Generalized hirsch h-index for disclosing latent facts in citation networks. *Scientometrics, 72*(2), 253–280.
Sonnenwald, D. (2007). Scientific collaboration: A synthesis of challenges and strategies. *Annual Review of Information Science and Technology* (41), 643–681.
Stokols, D., Harvey, R., Gress, J., Fuqua, J., & Phillips, K. (2005). In vivo studies of transdisciplinary scientific collaboration: Lessons learned and implications for active living research. *American Journal of Preventive Medicine, 28*(2), 202–213.
Tol, R. (2008). A rational, successive g-index applied to economics departments in Ireland. *Journal of Informetrics, 2*(2), 149–155.
van Raan, A. (2004). Measuring science. Capita select a of current main issues. *Handbook of quantitative science and technology research. The use of publication and patent statistics in studies of S&T systems*), pp. 19–50.
Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy, 34*(10), 1608–1618.
Yan, E., & Ding, Y. (2011). Discovering author impact: A pagerank perspective. *Information Processing & Management, 47*(1), 125–134.
Yin, L.-c., Kretschmer, H., Hanneman, R. A., & Liu, Z.-y. (2006). Connection and stratification in research collaboration: An analysis of the collnet network. *Information Processing & Management, 42*(6), 1599–1613.