Journal of Informetrics 5 (2011) 594–607

Contents lists available at ScienceDirect

Journal of Informetrics

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

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

Alireza Abbasi [a], Jörn Altmann [b,*], Liaquat Hossain [a]

[a] Centre for Complex Systems Research, Faculty of Engineering and IT, the University of Sydney, Sydney, NSW 2006, Australia
[b] Technology Management, Economics, and Policy Program & Department of Industrial Engineering, College of Engineering, Seoul National University, Seoul 151-744, South Korea

ARTICLE INFO

Article history:
Received 27 January 2011
Received in revised form 31 May 2011
Accepted 31 May 2011

Keywords:
Collaboration
Citation-based research performance
g-index
Co-authorship networks
Social network analysis measures
Regression
Correlation

ABSTRACT

In this study, we develop a theoretical model based on social network theories and analytical methods for exploring collaboration (co-authorship) networks of scholars. We use measures from social network analysis (SNA) (i.e., normalized degree centrality, normalized closeness centrality, normalized betweenness centrality, normalized eigenvector centrality, average ties strength, and efficiency) for examining the effect of social networks on the (citation-based) performance of scholars in a given discipline (i.e., information systems). Results from our statistical analysis using a Poisson regression model suggest that research performance of scholars (*g*-index) is positively correlated with four SNA measures except for the normalized betweenness centrality and the normalized closeness centrality measures. Furthermore, it reveals that only normalized degree centrality, efficiency, and average ties strength have a positive significant influence on the *g*-index (as a performance measure). The normalized eigenvector centrality has a negative significant influence on the *g*-index. Based on these results, we can imply that scholars, who are connected to many distinct scholars, have a better citation-based performance (*g*-index) than scholars with fewer connections. Additionally, scholars with large average ties strengths (i.e., repeated co-authorships) show a better research performance than those with low tie strengths (e.g., single co-authorships with many different scholars). The results related to efficiency 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. The negative effect of the normalized eigenvector suggests that scholars should work with many students instead of other well-performing scholars. Consequently, we can state that the professional social network of researchers can be used to predict the future performance of researchers.

1. Introduction

Performance appraisal is an inevitable function of management at any level. It fosters the development progress. Consequently, within a research environment (i.e., universities and research institutes), there should also be a performance evaluation for academics. This evaluation of researchers, which should be based on the researcher's output (i.e., productivity), is not only needed for performance appraisal but also for faculty recruitment, governmental funding

* Corresponding author.
E-mail address: jorn.altmann@acm.org (J. Altmann).

1751-1577/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2011.05.007

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.

Thus, there is a need for measuring the output of universities and the output of their researchers. With respect to governmental funding, i.e., the allocation of funding to a specific project, it is important to choose the most appropriate scholars with the aim of maximizing the research output, cost savings, and resource utilization. Therefore, the main problem is the identification of the most suitable scientists, who can achieve the goals (Jiang, 2008).

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 appraisal in the research community, if the researcher publishes and these publications get cited. The number of citations qualifies the quantity of publications (Lehmann, Jackson, & Lautrup, 2006). Hirsch (2005) introduced the *h*-index as a simple measure that combines in a simple way the quantity of publications and the impact of publications (i.e., number of citations). The *h*-index is defined as follows: "A scientist has an *h*-index of *h*, if *h* of her *Np* papers have at least *h* citations each, and the other ($Np - h$) papers have at most *h* citations each" (Hirsch, 2005). In other words, a scholar with an index of *h* has published *h* papers, which have been cited at least *h* times. Furthermore, the *h*-index also became the basis for a wide range of new measures for individuals assessment (Altmann, Abbasi, & Hwang, 2009; Batista, Campiteli, & Kinouchi, 2006; Egghe, 2006; Jin, 2006; Sidiropoulos, Katsaros, & Manolopoulos, 2007; Tol, 2008) and groups assessment (Altmann et al., 2009; Braun, Glänzel, & Schubert, 2006; Prathap, 2006; Schubert, 2007; Tol, 2008), by extending the previously mentioned indices to groups of scholars.

One of the most famous and widely used and accepted extension of the *h*-index is the *g*-index. The *h*-index and the *g*-index are well-established and widely used by academic databases (e.g., *Web of Science*[1] and *Scopus*[2]) to measure the performance of scholars. The *g*-index has been introduced by Egghe (2006) to overcome the main shortcomings of the *h*-index, namely, ignoring the number of citations in excess of *h*. Given a set of articles ranked in decreasing order of the number of citations that they received, the *g*-index is the (unique) largest number such that the top *g* articles received (together) at least g^2 citations (Egghe, 2006). The *g*-index takes into account both quality and quantity of output of researchers (similar to the *h*-index) and inhabits the simplicity and feasibility of the *h*-index.

While we evaluate associations of different scientific performance measures (i.e., publications count, citations count, *h*-index, and *g*-index) with SNA measures in this paper, we use the *g*-index measure as the main surrogate for quantifying the performance of researchers in our multivariate Poisson regression model.

The scientific landscape has also seen 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 and his colleagues (Luukkonen, Persson, & Sivertsen, 1992; Luukkonen, Tijssen, Persson, & Sivertsen, 1993) as well as Wagner and Leydesdorff (2005). They state that, by jointly publishing papers, researchers show their knowledge sharing activities, which are an indication for knowledge creation. Stokols, Harvey, Gress, Fuqua, and Phillips (2005) show that 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. With respect to the number of new publications, empirical studies have been conducted by Lee and Bozeman as well as Duque et al. (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), Lee and Bozeman (2005) show that the total number of publications for US scientists is positively associated with the total number of collaborations.

Consequently, it has been noticed that "the rising awareness of collaborativeness in science has led to a sharpened focus on the collaboration issue" (Melin, 2000). Furthermore, scientific collaboration has even been called a "springboard for economic prosperity and sustainable development" (US Office of Science & Technology Policy, 2000). As most scientific output is a result of group work and most research projects are too large for an individual researcher to perform, 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 level of individual researchers but also at the level of countries, most governments are interested in enhancing the level of international collaborations through policies (Katz & Martin, 1997; van Raan, 2004).

Since scientific collaborations can be 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. Since social networks may span disciplinary, organizational, and national boundaries, social networks can influence collaborations in multiple ways (Sonnenwald, 2007).

Currently, however, it is not clear which collaboration data is useful for evaluating the academic community. Although there is a large set of potential collaboration data (e.g., joined conference organization, joined research proposal submissions, joined publications, joined conference attendance, and teacher–student relationships), which qualifies for being analyzed through appropriate network measures, we only consider joined publications in our study. For our analysis, we use publication information that is available on the Internet. However, to restrict the data collection effort, we only selected publication data of scholars of five information systems schools (iSchools). For the data collection, we used a Web-based tool (Abbasi &

[1] http://science.thomsonreuters.com/training/wos/.

[2] http://help.scopus.com/robo/projects/schelp/h_hirschgraph.htm.

Altmann, 2010). Based on the co-authorships of those publications, we construct the research collaboration network of these scholars. Nodes (actors) of the research collaboration network represent scholars. A link (tie) between two nodes represents a publication co-authorship relationship between those two scholars. Within the remainder of the paper, we will refer to nodes and links if we talk about networks in general. The terms actors and ties are used, if we talk about the collaboration activity of scholars.

By calculating measures of social network analysis (SNA) and researcher's citation-based performance index (*g*-index), we aim to find whether the position of a researcher within the collaboration (co-authorship) network correlates with the research performance of this researcher. In particular, we investigate the following research questions:

- Which measures of SNA can be used to evaluate the co-authorship-based research collaboration network of scholars?
- Is there a correlation between measures of SNA and scholars' citation-based performance measures?
- Which measures of SNA of scholars have an impact on the scholar's performance, in particular on *g*-index?
- What are the implications of our findings for scholars and research communities with respect to their productivity improvement?

The remainder of the paper is organized as follows: Based on a review of existing studies about social network analysis measures and performance of researchers, we introduce our hypotheses about the usefulness of SNA measures for evaluating research performance in Section 2. Section 3 describes the data resources, the research methodology applied for the data collection and validation, and our analysis model. Section 4 analyzes the collaboration network of five iSchools. In particular, it presents the results of the Spearman rank correlation test between SNA measures and the performance measure and also shows the Poisson multiple regression result on the influence of SNA measures on research performance. Finally, we conclude the paper with a discussion on the results, the research limitations, and our future work.

2. Social network analysis measures and theories

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, 2010; Kim & Altmann, 2010). 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 a efficient way and to evaluate the performance of individuals, groups, or the entire social network (Abbasi & Altmann, 2010). For instance, with respect to performance evaluation, the social network of a researcher within a research community provides an indication of the researcher's collaboration activity (Abbasi, Altmann, & Hwang, 2010).

2.1. Network structures

The "Bavelas–Leavitt Experiment" is one of the earliest studies that relates human communication patterns to performance (Bavelas, 1947, 1950; Leavitt, 1951). The experiment consisted of several groups by five members, who had to communicate with each other through enclosed cubicles to solve a puzzle. Several different structures for communication channels between members of the groups have been found as shown in Fig. 1. Their result showed better performance for the groups using a "Star" and "Y" structure. They inferred that centralization was the most influential factor on performance (Leavitt, 1951). However, other studies showed that it is only true for simple, standard and routine tasks (Chung & Hossain, 2009). Guetzkow and Simon (1955) found that decentralized structures (e.g., circle network) lead to efficient performance when solving complex tasks.

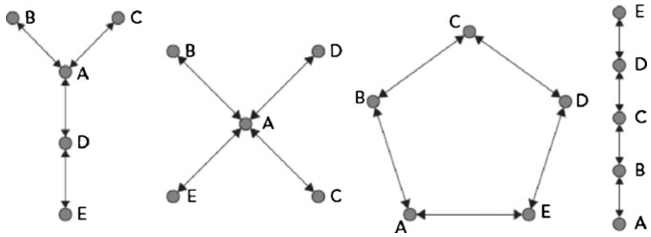

Fig. 1. The Y, Star, Circle and Line structures of communication. From Chung and Hossain (2009).

2.2. Centrality measures and theories

Another method used to understand networks and their participants is to evaluate the location of nodes in the network. Measuring the network location is about determining the centrality of a node. These measures help determining the importance of a node in the network. Bavelas (1950) has been the pioneer, who initially investigated the formal properties of centrality and proposed several centrality concepts. Later, Freeman (1979) found that centrality has an important structural influence on leadership, satisfaction, and efficiency. In particular, it could be shown that betweenness centrality and degree centrality influence the performance of a node.

In addition to this, a node can be central from a local or global perspective. A node is locally central, if it has a large direct neighborhood of nodes. It is important to recognize that this property does not mean that the node is a unique central node in the network. If a node is globally central, it has a position of strategic significance in the overall structure of the network (Scott, 1991).

2.2.1. Degree centrality

The simplest and easiest way of measuring node centrality is by measuring the degree of the node in the graph. The degree of a node is simply the number of other nodes connected directly to the node. As degree of a node is calculated in terms of the number of its adjacent nodes, the degree can be regarded as a measure of local centrality (Scott, 1991). Thus, the degree centrality of node p_k is given by:

$$C_D(p_k) = \sum_{i=1}^{n} a(p_i, p_k) \tag{1}$$

where n is the number of nodes in the network and $a(p_i,p_k)$ is a distance function. $a(p_i,p_k) = 1$, if and only if node p_i and node p_k are connected. $a(p_i,p_k) = 0$ otherwise.

It is not meaningful to compare a node with a score of 40 in a network of 100 nodes with a node of score of 7 in a network by 10 nodes. In order to have a more general measure for comparing the degree centrality of nodes of different networks with different sizes, Freeman (1979) proposed a relative (normalized) measure. This measure normalizes the actual number of links by the maximum number of links it could have. Thus, the normal (relative) degree centrality of node p_k is given by:

$$C_{D'}(p_k) = \frac{\sum_{i=1}^{n} a(p_i, p_k)}{n-1} \tag{2}$$

Having just regular degree centrality measures, we can only compare nodes in networks with the same size. The normalized centrality measure, however, makes possible the comparison of node centrality across networks with different sizes.

In practice, an actor with a high degree centrality can influence a group by withholding or distorting information in transmission (Bavelas, 1950; Chung & Hossain, 2009; Freeman, 1979; Leavitt, 1951). The degree centrality is also an indicator of an actor's communication activity or popularity.

2.2.2. Closeness centrality

Local centrality measures are expressed in terms of the number of nodes to which a node is connected (Scott, 1991) but Freeman (1979, 1980) proposed *closeness* as a measure of global centrality in terms of the *distance* among various nodes. Therefore, a node is globally central, if it lies in average at the shortest distance from all other nodes. That means, it is 'close' to all other nodes in the network. Sabidussi (1966) used the same concept in his work as 'sum distance', the sum of the 'geodesic' distances (the shortest path between any particular pair of nodes in a network) to all other nodes in the network. Freeman (1979, 1980) defined closeness of a node as the "*sum of reciprocal distance*" of that node to any other nodes. So, closeness centrally of node p_k is given by:

$$C_C(p_k) = \sum_{i=1}^{n} d(p_i, p_k)^{-1} \tag{3}$$

In order to use this measure for comparing nodes across networks with different sizes, there is a need to normalize this measure. The measure can be normalized by using the maximum possible distance between any two nodes in a network of n nodes. This value is $n - 1$. More precisely, the normalized closeness of node p_k is given by:

$$C_{C'}(p_k) = \frac{\sum_{i=1}^{n} d(p_i, p_k)^{-1}}{n-1} \tag{4}$$

A node, which is in the on average nearest position to all other nodes, can most efficiently obtain information. Therefore, closeness is a surrogate measure for the independence and efficiency for communicating with other nodes in the network (Freeman, 1979).

2.2.3. Betweenness centrality

Freeman (1979) proposed another concept of node centrality, which measures the number of times a particular node lies 'between' the various other nodes in the network. This measure, which is called betweenness centrality, is defined as "the number of shortest paths (between all pairs of nodes) that pass through a given node" (Borgatti, 1995). Therefore, the betweenness of node p_k is given by:

$$C_B(p_k) = \sum_{i<}^{n} \sum_{j}^{n} \frac{g_{ij}(p_k)}{g_{ij}}, \quad i \neq j \neq k \tag{5}$$

where g_{ij} is the number of geodesics (shortest paths) linking node p_i and node p_j and $g_{ij}(p_k)$ is the number of geodesics linking node p_i and node p_j that contains node p_k.

In the same way, to be able to use this measure for comparing nodes across different networks with different sizes, there is a need to normalize (standardize) this measure. The measure is normalized by the maximum possible number of shortest paths (excluding the node under consideration). Given that the network is undirected, the maximum is: $((n-1)(n-2))/2 = (n^2 - 3n + 2)/2$. Therefore, the normal (relative) betweenness centrality score of node p_k is given by:

$$C'_B(p_k) = \frac{2 * C_B(p_k)}{n^2 - 3n + 2} \tag{6}$$

Betweenness is an indicator of the potential of a node (actor), which plays the role of a broker or gatekeeper. It can most frequently control information flows in the network (Freeman, 1979).

2.2.4. Eigenvector centrality

Based on the idea that an node is more central if it is linked to nodes that are themselves central (Bonacich, 1972), it is argued that the centrality of a node does not only depend on the number of its adjacent nodes but also on the values of centrality of these adjacent nodes. A node, which is connected to many other nodes that are themselves well-connected, has a high eigenvector centrality and a node connected to nodes with a few connections has a much lower score (Lu, Luo, Polgar, & Cao, 2010). Therefore, Bonacich (1972) defines the centrality $c(p_k)$ of a node p_k as positive multiple of the sum of adjacent centralities, i.e.:

$$\lambda * C_E(p_k) = \sum_{k=1}^{n} (a_{ik} * C_E(p_k)) \quad \forall i \tag{7}$$

Considering the centrality of all nodes and representing $c = (c(v_1), \ldots, c(v_n))$, the set of formulas can be written in matrix notation as $\lambda c = Ac$. This type of equation is well-known and can be solved by calculating the eigenvalues and eigenvectors of A. As Bonacich (1972) shows, only one eigenvector of the set of resulting eigenvectors is an appropriate solution that can serve as a centrality measure. As A is the adjacency-matrix of an undirected (connected) graph, A is non-negative and, due to the theorem of Perron–Frobenius, there is an eigenvector of the maximal eigenvalue with only non-negative (positive) entries (Ruhnau, 2000).

In the same way as for the other measures, there is a normalized version of the eigenvector measure. We use the Euclidean-norm for normalizing the eigenvector centrality:

$$C'_E(p_k) = \frac{C_E(p_k)}{\sqrt{\sum_{i=1}^{n} C_E(p_i)^2}} \tag{8}$$

The normalized eigenvector centrality is in the range $[0,(0.5)^{0.5}]$. This definition of normalized eigenvector centrality is not to be confused with the definition used in UCINET by Borgatti, Everett, and Freeman (2002). Furthermore, the normalized eigenvector centrality can be scaled "by the square root of one half, which is the maximum score attainable in any graph" (Borgatti & Everett, 1997):

$$C_E(p_k) = \sqrt{2} C'_E(p_k) \tag{9}$$

For our analysis, we consider the normalized eigenvector centrality $C''_E(p_k)$.

2.2.5. Centrality-related hypotheses

In line with these arguments, it is expected that authors have a high potential to have good performance, if they have many collaborations (links), are the closest authors to all other authors, are on many geodesic paths between other pairs of authors, and are connected with other centrally located authors. That means good-performing authors are in the center of a collaboration network. As the performance measure of scholars, as explained earlier, the *g*-index is used. Therefore, the following hypotheses are formally derived as:

- **H1a**: Normalized degree centrality of a scholar impacts her research performance (e.g., *g*-index).
- **H1b**: Normalized closeness centrality of a scholar impacts her research performance (e.g., *g*-index).
- **H1c**: Normalized betweenness centrality of a scholar impacts her research performance (e.g., *g*-index).
- **H1d**: Normalized eigenvector centrality of a scholar impacts her research performance (e.g., *g*-index).

2.3. Tie strengths theories

Another point of view to analyze actors of a network has been introduced by Granovetter (1973). He established the theory of the 'strength of weak ties', which argues that an individual obtains new and novel information from weak ties rather than from strong ties within the individual's group structure. It is because new information originates via weak ties, which serve as bridges to different clusters of people (Chung & Hossain, 2009). Granovetter (1973) defined strength of a tie as "a combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie".

Contrarily, Krackhardt (1992) showed that strong ties are important in the generation of trust. He introduced the theory of 'strength of strong ties' in contrast to Granovetter's (1973) theory. Levin and Cross (2004) found that strong ties, more so than weak ties, lead to the receipt of useful knowledge for improving performance in knowledge-intensive work areas. However, controlled for the dimension of trust, the structural benefit of weak ties emerged in their research model. It suggests that the weak ties provide access to non-redundant information. Weak ties facilitate faster project completion times, if the project is simple. It enables faster search for useful knowledge among other organizational subunits. Strong ties foster complex knowledge transfer, if knowledge is highly complex (Hansen, 1999; Reagans & Zuckerman, 2001).

For a weighted network (graph), the strength of a tie (link between two nodes) simply is the weight of the link. In our model, ties are weighted by the number of collaborations between two co-authors.

To evaluate a node's ties strength, we calculate the average of the weights of his co-authorships (ties). That means we divide the sum of the node's tie weights (i.e., the number of collaborations of the scholar) by the degree of the node (i.e., the scholar's total number of different co-authors). The average ties strength TS of node p_k is given by:

$$\mathrm{TS}(p_k) = \sum_{i=1}^{n} \frac{W_{ki}}{C_D(p_k)} \tag{10}$$

where W_{ki} represents the weights of the ties between node p_k and node p_i and $C_D(p_k)$ represents degree centrality of node p_k.

Thus, based on these definitions of strength of ties, the following hypothesis is proposed:

H2 Average ties strength of a scholar impacts her research performance (e.g., *g*-index).

2.4. Structural holes theory and efficiency

Freeman's (1979) approach to betweenness is build around the concept of 'local dependency'. A node p_i is dependent upon another node p_j, if paths which connect it to other nodes pass through node p_j (Scott, 1991). Burt (1995) has described this in terms of 'structural holes' and made an influential contribution to the phenomena of structural effects on individual outcome by relating the theory of structural holes to network structure and network position. A structural hole exists where two nodes are connected at a distance of 2 but not at distance of 1. The existence of a structural hole allows the third node (i.e., the node which is between the two nodes) to act as a broker or intermediary (Scott, 1991). In other words, holes in the network refer to the absence of links that would otherwise connect unconnected clusters together. Individuals, who bridge these holes, attain an advantageous position that yields information and control benefits (Burt, 1995). Structural holes theory is based on betweenness centrality.

Burt uses the theory of structural holes to optimize a network. Burt claims that increasing the number of direct contacts (ego-network size) without considering the diversity reached by the contacts makes the network inefficient in many ways (Burt, 1995; Chung, 2009). Therefore, the number of non-redundant contacts (i.e., the four nodes in Network B of Fig. 2) is important to the extent that redundant contacts would lead to the same people and, hence, provide the same information and control benefits. Besides, the Network A of Fig. 2 is inefficient as the node ("ego") gets only redundant information from

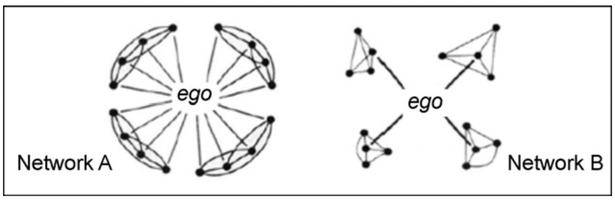

Fig. 2. Two networks with structural holes, while network A is less efficient than network B. Adapted from Burt (1995).

its primary contacts. This node is wasting its resources to maintain its ties to all actors of the same cluster, which usually spread the same information. Non-redundant collaborators, however, give access to a diversity of information, which usually leads to innovation and high performance.

Since this definition of efficiency of a node appears to be helpful in the context of our research collaboration network, we follow the definitions of Burt (1995, 2009). In detail, efficiency is defined as the ratio of the total number of disjoint groups of primary nodes (i.e., neighbors) of node p_k, where the nodes of such a group are only connected to nodes of the same group but not to nodes of other groups, and the number of primary nodes of node p_k (i.e., the degree centrality of node p_k). Thus, efficiency of node p_k is given by:

$$E(p_k) = \frac{g(p_k)}{C_D(p_k)} \tag{11}$$

where $g(p_k)$ denotes the number of disjoint groups of primary contacts of p_k and $C_D(p_k)$ represents degree centrality of node p_k.

With respect to our study, a disjoint group of primary contacts relates to co-authors that have joined publications. 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. Therefore, in order to test this property, we formulate the following hypothesis:

H3 The scholar's efficient use of co-authors impacts her research performance (e.g., *g*-index).

3. Data and methodology

3.1. Data collection

For this study, we collected data on the information schools (iSchools) of five universities: University of Pittsburgh, University of Berkley, University of Maryland, University of Michigan, and Syracuse University. These schools have been chosen, since they offer similar programs in the area of information management and systems, and since the topic of these schools is new within the university landscape.

The data sources used are the school reports, which include the list of publications of their scholars, DBLP (http://www.informatik.uni-trier.de/~ley/db), Google Scholar (http://scholar.google.com), and ACM portal (http://portal.acm.org). Citation data has been taken from Google Scholar and ACM Portal, using AcaSoNet (Abbasi & Altmann, 2010). AcaSoNet is a Web-based application for extracting publication information (e.g., author names, title, publication date, publisher, and number of citations) from the Web. It also extracts relationships (e.g., co-authorships) between researchers and stores the data in the format of tables in its local database.

For its citation counting service, Google Scholar considers a variety of publication databases, which belong to different publishers and list different types of publications. Thus, it produces a higher publication count per researcher and a higher citation count per publication than other citation counting services (e.g., Web of Science of Thomson Reuters, and Scopus) (Kousha & Thelwall, 2007). Consequently, the calculation of the *h*-index and the *g*-index, if based on Google Scholar, results in higher values than for the other citation counting services. However, Ruane and Tol (2008) showed that rankings based on Google Scholar have a high rank correlation with rankings based on Web of Science or Scopus.

For our analysis, we followed Google Scholars approach and did not differentiate between the different types of publications (e.g., proceedings of local conferences, proceedings of international conferences, journals, books, and presentations were weighted equally). Our data covered a period of 5 years (2001–2005), except for the University of Maryland iSchool, which had no data for the year 2002 in their report. To resolve this issue, we substituted the missing data with data of the year 2006. As we do not apply longitudinal analysis, this does not constitute a problem.

Despite AcaSoNet, much data cleansing has become necessary in order to allow processing of the extracted publication data. Most of the cleansing was due to the lack of a single standard format used for listing publications (e.g., the order of first name and family name of authors, the order of title and publication year and the inaccuracy in writing journal and conference names). After the cleansing of the publication data of the five iSchools, data about 2139 publications, 1806 authors, and 5310 co-authorships was finally available for our analysis. Table 1 shows a characterization of the co-authorship network with respect to the five iSchools. In particular, it highlights the number of professors and the number of authors within the different iSchools.

Table 1
iSchools' network measures.

	Pittsburgh	Maryland	Michigan	Syracuse	Berkeley
Number of authors	358	303	603	280	262
Number of authors, who are professors	26	13	44	33	11
Number of publications (Output)	477	312	490	375	468

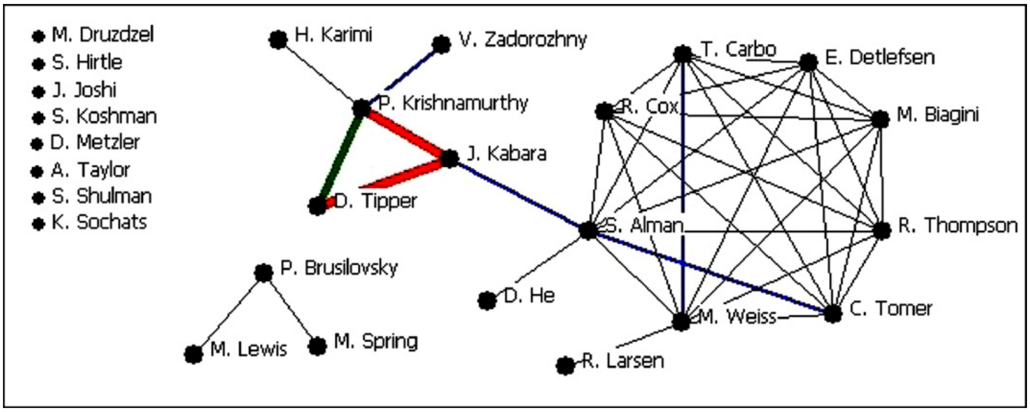

Fig. 3. An example of a co-authorship network. The co-authorship network of Pittsburgh iSchools faculty members.

3.2. Methodology

Social networks are represented as graphs, which are constructed of nodes (actors) and links (ties). Nodes, which denote individuals, organizations, or information, are linked, if one or more specific types of ties (e.g., financial exchange, friendship, trade, and Web links) exist between them. For example, a node could represent a person and a link between two nodes could represent that these two persons know each other.

The co-authorship network (i.e., the research collaboration network) is represented through a graph. The nodes (actors, participants, vertices) i of the graph represent researchers (scholars). A link (tie, relation, edge) a_{ij} between node i and node j indicates a collaboration relationship between nodes, based on the co-authorship of researchers on publications. Publications, of which the author is the sole author, are presented through loops (i.e., a link from a node to itself) in the graph. The weight of a link w_{ij} denotes the number of publications that two researchers co-authored. Fig. 3 shows the internal co-authorship network of the Pittsburgh iSchools' faculty as an example. Different link weights are indicated through different colors and link strengths.

After preparing the social network matrix, we used UCINET (Borgatti et al., 2002) as a tool for visualizing the network and for calculating the network measures (i.e., normalized degree centrality, normalized betweenness centrality, normalized closeness centrality, normalized eigenvector centrality, ties strength, and efficiency of each node of the co-authorship network (i.e., the research collaboration network).

In the next step, after calculating the scholar's network measures and the g-index, we use the Spearman rank correlation test to test our hypotheses (Fig. 4) and to find the associations between independent variables and dependant variable.

Finally, to find the predicators of the research performance measure, we use a Poisson multiple regression model. We use Poisson regression model, since the observations of g-index values can be assumed to be Poisson distributed. Besides, the maximum likelihood method used can scope with potential multi-collinearity between the independent variables. A Poisson regression model is suited as our dependent variable (g-index) is a simple transformation of the citation count variable, is not over-dispersed, and does not have an excessive number of zeros (about 15%).

The results of the analysis of the five co-authorship networks, representing the collaborations of researchers of each of the five iSchools, is shown in this paper and discussed.

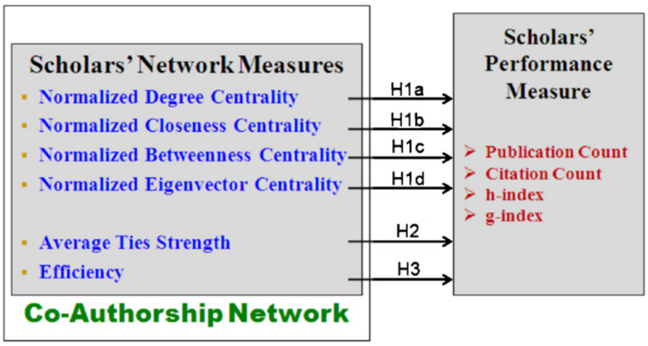

Fig. 4. Research model to investigate the effect of scholar's network measures on the performance of scholars.

4. Analysis and results

Using our data, we calculate the SNA measures of our model and the *g*-index for all scholars of all five iSchools. The results for the top 30 performing scholars are shown in Table 2.

4.1. Spearman rank correlation test

We test our hypotheses based on the data of Table 2 but for all researchers. For this, the Spearman rank correlation test is used. In particular, we calculate the Spearman rank correlation coefficient between the six SNA measures and four performance measures (i.e., the publication count, the citation count, the *h*-index, and the *g*-index). The results are shown in Table 3.

As the results show, the correlation coefficients between the SNA measures (i.e., normalized degree centrality, normalized betweenness centrality, efficiency, and average ties strength) and the performance measures are high. An exception is

Table 2
Name, normalized degree centrality, normalized closeness centrality, normalized betweenness centrality, normalized eigenvector centralities, efficiency, average ties strength, and *g*-index of the top 30 performing researchers.

	Name	Normalized degree centrality	Normalized closeness centrality	Normalized betweenness centrality	Normalized eigenvector centrality	Efficiency	Average ties strength	*g*-Index
1	Peter Brusilovsky	0.090	0.011	0.252	0.014	0.987	2.063	37
2	Marti Hearst	0.091	0.006	0.037	0.001	0.912	2.522	33
3	Martha E. Pollack	0.065	0.008	0.071	0.000	0.942	1.795	31
4	Elliot Soloway	0.057	0.008	0.031	0.000	0.783	3.059	30
5	Kevin Crowston	0.101	0.025	0.168	0.015	0.935	2.679	30
6	Jimmy Lin	0.161	0.020	0.152	0.225	0.896	2.204	26
7	Dragomir R. Radev	0.152	0.008	0.187	0.250	0.869	1.527	26
8	Hal Varian	0.044	0.006	0.016	0.000	0.949	1.636	25
9	Douglas W. Oard	0.306	0.020	0.347	0.558	0.927	1.935	24
10	Steven P. Abney	0.117	0.008	0.413	0.000	0.955	1.386	24
11	Allison Druin	0.188	0.020	0.192	0.014	0.843	2.211	23
12	John Chuang	0.111	0.008	0.088	0.000	0.935	1.964	22
13	Danah boyd	0.000	0.000	0.000	0.000	0	0	21
14	Mark S. Ackerman	0.057	0.008	0.119	0.000	0.952	1.471	21
15	Edmund H. Durfee	0.077	0.008	0.107	0.000	0.877	1.565	21
16	Mark W. Newman	0.040	0.008	0.030	0.000	0.724	2.958	21
17	Judith S. Olson	0.047	0.008	0.075	0.000	0.836	2.143	21
18	Ping Zhang	0.083	0.025	0.113	0.040	0.93	1.957	21
19	Clifford Lynch	0.024	0.004	0.000	0.000	0.667	1	20
20	Anna Lee Saxenian	0.012	0.004	0.000	0.000	0.778	1	20
21	Jennifer J. Preece	0.141	0.020	0.173	0.006	0.933	1.395	20
22	Joseph Krajcik	0.048	0.008	0.015	0.000	0.705	2.414	20
23	Richard J. Cox	0.056	0.011	0.042	0.000	0.817	1.05	19
24	Gary M. Olson	0.043	0.008	0.065	0.000	0.826	1.846	19
25	Elizabeth D. Liddy	0.140	0.025	0.207	0.559	0.884	2.179	19
26	Michael Lewis	0.101	0.011	0.174	0.001	0.861	3.306	18
27	Chris Quintana	0.033	0.008	0.008	0.000	0.555	2.95	18
28	John Canny	0.103	0.008	0.077	0.000	0.911	1.192	17
29	John L. King	0.032	0.002	0.002	0.000	0.963	1.316	17
30	Joon S. Park	0.076	0.025	0.136	0.007	0.98	1.476	17

Table 3
Spearman rank correlation test for scholars of five iSchools, showing the correlation value and the significance level ($N = 1809$) for the citation-based performance measures and the SNA measures.

	Variables	1	2	3	4	5	6	7	8	9
1	Publication Count	–								
2	Citation Count	0.516**	–							
3	*h*-index	0.762**	0.780**	–						
4	*g*-index	0.820**	0.765**	0.964**	–					
5	Normalized degree centrality	0.288**	0.332**	0.311**	0.305**	–				
6	Normalized closeness centrality	0.024	−0.012	0.052*	0.055*	0.247**	–			
7	Normalized betweenness centrality	0.585**	0.388**	0.501**	0.529**	0.406**	0.162**	–		
8	Normalized eigenvector centrality	0.000	0.060*	0.041	0.041	0.411**	0.533**	0.106**	–	
9	Efficiency	0.384**	0.120**	0.281**	0.308**	−0.354**	0.080**	0.352**	−0.125**	–
10	Average ties strength	0.861**	0.434**	0.660**	0.701**	0.229**	0.088**	0.395**	0.016	0.345**

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 4
Omnibus test of Poisson multiple regression for six independent variables.

Likelihood ratio chi-square	df	Sig.
3061.315	6	0.000

only the normalized closeness centrality and the normalized eigenvector centrality. The normalized closeness centrality measure of researchers does not correlate to scholars' publication count and citation count. It shows also only a very little but significant correlations with the *h*-index and the *g*-index. Besides, the eigenvector centrality measure correlates only very little with the citation count.

Considering the *g*-index measure as the main indicator of scholars' performance, the Spearman rank correlation coefficient shows a significant positive correlation with four of the six SNA measures (i.e., normalized degree centrality, betweenness centrality, average ties strength, and efficiency). The *g*-index is not correlated with the normalized eigenvector centrality measure and shows only a very little correlation with the normalized closeness centrality measure. Besides, the significance level and the correlation for ties strength and betweenness centrality are higher than for the efficiency measure and the degree centrality measure.

Based on these results, we can state that there is correlation between four SNA measures (i.e., normalized degree centrality, normalized betweenness centrality, efficiency, and average ties strength) and the research performance (i.e., publication count, citation count, *h*-index, and *g*-index). Positively significant correlations expose that researchers, who are more central (i.e., having many collaborations with different scholars; being frequently between the collaboration paths of other scholars; maintaining collaborations with one scholar of a group of collaborating scholars) are more productive. In addition to this, scholars, who have strong ties (i.e., repeated co-authorships) to co-authors, have a better research performance than those with low ties (e.g., single co-authorships with many different co-authors). Therefore, the theory of 'Strength of Strong Ties' by Krackhardt (1992), which has been explained in an earlier section is supported by our analysis. Besides, the positive correlation between the performance measures and efficiency shows that researchers have to be selective about the ties that they maintain, following Burt (1995). Thus, non-redundant co-authorship relationships (i.e., maintaining strong relationships to only one co-author of a group of linked co-authors (highly connected together)) will result in improved performance of the scholar.

Compared with the result shown in Abbasi and Altmann (2011), it becomes obvious that the data set used has a tremendous impact on the correlation results. While in the previous study only professors had been considered, this study comprises professors, students and collaborating researchers of professors at the five iSchools. As Table 1 shows, the number of authors of an iSchool is manifold higher than the number of professors. Professors, who have a higher performance measure than students, have a brokering role, connecting their students to other researchers. Consequently, the normalized betweenness centrality in this study gets higher correlation with the performance measures. At the same time, because of the many relationships of professors with their students, the eigenvector centrality measure of the professor is low because of the many surrounding authors (students), which also have a low eigenvector centrality measure.

4.2. Poisson multiple regression

Since our Spearman rank correlation analysis shows only the existence of relationships between SNA measures and performance measures but not the effect of independent variables on dependent variables, we use a multivariate regression model (multiple regression analysis). Gibbons (1982) has been the first to suggest a multivariate regression model (MVRM) methodology to measure the effect of new information on asset prices. Later Binder (1985) showed advantages of the MVRM methodology over other event study methodologies. With our MVRM, we test the effects of co-authorship network measures (independent variables) on the author's performance measures (dependent variable).

In particular, we run a Poisson multiple regression model, in order to identify which of the independent variables (SNA measures) impacts the dependent variable (*g*-index). The Poisson regression models the log of the expected count as a function of the predictor variables. For this regression, we executed the Poisson regression with the robust option (i.e., to get robust standard errors for the Poisson regression coefficients) in our statistical software (SPSS).

As Table 4 indicates, comparing the fitted model against the intercept-only model, the model is significant ($p = 0.000$). The result of the multiple regression model show (Table 5) that the significant variables are normalized degree centrality ($\beta = 16.799$, $p = 0.000$), normalized eigenvector centrality ($\beta = -3.441$, $p = 0.001$), efficiency ($\beta = 1.157$, $p = 0.000$) and average ties strength ($\beta = 0.386$, $p = 0.000$). The β values are the estimated Poisson regression coefficients for the model. As expected from the results of the Spearman rank correlation, the normalized closeness centrality measure is not significant in this regression. As a surprise, at the first glance, comes that the normalized betweenness centrality coefficient is not significant but the normalized eigenvector centrality coefficient (which is even negative). At the second glance, it becomes clear that the ranking, which is based on the betweenness centrality measure, positions professors higher than students. The absolute values of the between centrality measure, however, do not differ largely within the network. With respect to the normalized eigenvector centrality measure, it can be stated that a professor is more successful if the professor has many students around instead of successful colleagues.

Table 5
Poisson multiple regression results for six independent variables and the g-index as dependent variable.

Parameters	β	Std. error	Hypothesis Test		
			Wald Chi-Square	df	Sig.
Intercept	−0.652	0.1207	29.161	1	0.000
Normalized degree centrality	16.799	2.4134	48.455	1	0.000
Normalized Closeness Centrality	−8.023	4.0851	3.857	1	0.050
Normalized Betweenness Centrality	0.867	0.9890	0.769	1	0.381
Normalized eigenvector centrality	−3.441	1.0782	10.184	1	0.001
Efficiency	1.157	0.1171	97.593	1	0.000
Average ties strength	0.386	0.0559	47.754	1	0.000
Scale	1[a]				

[a] Fixed at the displayed value.

Table 6
Omnibus test of Poisson multiple regression four independent variables.

Likelihood ratio chi-square	df	Sig.
3045.084	4	0.000

Based on this regression result (i.e., low significance of the normalized closeness centrality, and the normalized betweenness centrality), we execute a second regression with four independent variables only: normalized degree centrality, normalized eigenvector centrality, efficiency, and average ties strength. The omnibus test for this regression shows also its significance ($p = 0.000$) against the intercept-only model (Table 6). All independent variables are significant ($p = 0.000$). It is to note that the coefficients of average ties strength, efficiency, normalized eigenvector centrality, and normalized degree centrality only changed slightly, compared to the previous regression. This illustrates the stability of the model. The high Wald chi-square values, shown in Table 7, suggest that the coefficients describe the log(g-index) well.

Therefore, taking the results of our regression, the regression equation for our analysis can be written as:

$$\text{Log}_e(g\text{-index}) = -0.762 + (17.681 * CD) - (3.587 * CE) + (1.181 * E) + (0.386 * TS) \tag{12}$$

This can be interpreted as, having one more publications with a new researcher in a collaboration network of 100 researchers (i.e., increasing normalized degree centrality by 0.01) will increase the g-index by approximately one unit ($1.193 = e^{(17.681*0.01)}$), assuming all other variables constant. In reality though, adding one co-author will not change one centrality measure but all centrality measures.

4.3. Discussion of hypotheses

Based on our analysis, we can state that the two independent variables, normalized closeness centrality and normalized betweenness centrality, do not show a significant impact on the g-index performance measure. Although the normalized betweenness centrality showed a positive Spearman rank correlation, even this independent variable was not a significant predictor in the regression model. The reason is the dominance of the high-performing professors. They have many students around them in the network. This causes to rank them higher than the students with respect to the normalized betweenness centrality and, at the same time, gives them a low normalized eigenvector centrality rank (because of the low values). Thus, we can only accept hypotheses H1a, H1d, H2 and H3 and infer that:

- **H1a**: Normalized degree centrality of a researcher has a (positive) impact on her research performance (g-index).
- **H1d**: Normalized eigenvector centrality of a researcher has a (negative) impact on her research performance (g-index).
- **H2**: Average ties strength of a researcher has a (positive) impact on her research performance (g-index).
- **H3**: Efficiency of a researcher has a (positive) impact on her research performance (g-index).

Table 7
Poisson multiple regression results for four independent variables.

Parameters	β	Std. error	Hypothesis test		
			Wald chi-square	df	Sig.
Intercept	−0.762	0.1093	48.556	1	0.000
Normalized degree centrality	17.681	1.8110	95.311	1	0.000
Normalized eigenvector centrality	−3.587	1.0252	12.241	1	0.000
Efficiency	1.181	0.1142	106.923	1	0.000
Average ties strength	0.386	0.0552	48.896	1	0.000
Scale	1[a]				

[a] Fixed at the displayed value.

Consequently, we can state that the researcher performance will improve, if the researcher has many distinct co-authors, has repeated collaborations with each of her co-authors, and is connected to a single researcher of a disjoint group of researchers. That means, scholars should not only keep strong relationships with existing co-authors and build on former co-authorships but also try to have collaborations with new authors. However, in order to increase the efficiency, scholars should only keep strong relationships with one co-author of a group of linked researchers.

On the other hand, being close to other researchers (in the average nearest position to all other authors) and also having a brokerage position among researchers in a co-authorship network does not have an effect on the performance of the researcher. It suggests that there is a high chance of receiving redundant information from that group of connected researchers which may lead to a low performance.

Besides, as the normalized eigenvector centrality reflects a researcher's connections to other well-connected people (Lu et al., 2010), our results suggest that the performance can be increases by collaborating with many students or researchers that have a low performance record at the time of the collaboration.

5. Discussion and conclusion

In order to improve the benefit from research (and research funding), well-performing researchers, who can manage and control a scientific research group, have to be identified. As past research has shown, the *g*-index can be a surrogate for evaluating the research performance of scholars. To measure the collaboration skills of researchers, which became more and more important for research management over the past years, the co-authorships are often used. Consequently, we considered both in this study.

In particular, in order to investigate whether the collaboration skills are correlated with and have an impact on the research performance of researchers, we used a co-authorship data set of professors and students of 5 iSchools. The co-authorship data is used to derive the collaboration network of researchers. The analysis of the collaboration network is performed by applying social network analysis measures. The social network analysis measures used are: the normalized degree centrality, the normalized closeness centrality, the normalized betweenness centrality, the normalized eigenvector centrality, the average ties strength, and the efficiency. Our analysis comprised a Spearman rank correlation analysis and a regression analysis.

The results of our Spearman rank correlation analysis show that the research performance is positively associated with all social network measures. However, the coefficient correlations for normalized degree centrality and normalized betweenness centrality, average ties strength, and efficiency are the highest and significant. With respect to the normalized degree centrality, scholars, who are connected to many different scholars, show better performance than those with fewer connections. Scholars with strong ties (i.e., repeated co-authorships) show a better research performance than those with low ties (e.g., single co-authorships with many co-authors). 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 of a group of linked co-authors. Therefore, we can state that scholars should avoid collaboration with authors within the same cluster. It would lead to lower efficiency.

Performing a multiple Poisson regression analysis for identifying which social network measure influences the performance of scholars, we found that normalized degree centrality, normalized eigenvector centrality, average ties strength, and efficiency are correlated with scholars' performance (i.e., the *g*-index). However, only the normalized degree centrality, the average ties strength, and the efficiency have positive effects on scholars' citation-based performance (*g*-index) and the normalized eigenvector centrality has a negative impact on the *g*-index. These results indicate that scholars, who are locally central (i.e., have many and strong direct contacts (co-authors) though restricted to co-authors that are the only access to a group of linked researchers) perform better than the ones who are globally central in the network. Degree centrality is a surrogate for local centrality, while closeness centrality and betweenness centrality show how central a node is globally. We can explain the lack of effectiveness of closeness and betweenness centralities by discussing the frequency of knowledge exchange between scholars and their direct contacts (co-authors) rather than indirect contacts (co-authors of co-authors). In other words, the number of co-authors of a scholar determines the opportunities to collaborate and exchange knowledge, leading to improved performance. Occupying a central position in a network in terms of closeness and/or betweenness gives only strategic importance to the scholar but does not necessarily improve her performance. Therefore, having short paths (i.e., high closeness centrality) to scholars, who do not have a direct co-authorship relationship but can be reached via a co-authorship path, might cause the exchange of redundant knowledge.

The normalized eigenvector centrality considers the co-authors' centrality. Although the low Spearman rank correlation and the strong negative correlation seem to be contradicting at a first glance, they fit considering the data set mixed of professors and students. It is caused by the fact that well-performing professors supervise many students, which have a low eigenvector centrality. Consequently, the normalized eigenvector centrality measure of the well-performing professor is low because of the many students with even lower eigenvector centralities. At the same time, it increases the normalized betweenness centrality since the performing professor provides the only connectivity for her students to other co-authors. Consequently, the negative regression coefficient for the normalized eigenvector centrality shows that the well-performing professor is successful because of this supervision of many students instead of being connected with other well-performing researchers.

Finally, access to demographic information of researchers (e.g., age, gender, and nationality) could be useful as moderating variables in our model. We would be able to categorize researchers and analyze the outcome for each of the categories. It could help us finding a generalization of our model. The current lack of access to this kind of information can be considered a limitation of our research.

Additional social context information of authors such as the role of authors (e.g., advisor, student, and colleague) could be useful to extend this research to perform a student-centric study of scientific collaboration networks following the research of Suresh, Raghupathy, Shekar, and Madhavan (2007). It would allow studying the dynamics of collaborations between students and professors.

References

Abbasi, A., & Altmann, J. (2011). On the correlation between research performance and social network analysis measures applied to research collaboration networks. In *Hawaii International Conference on System Sciences, Proceedings of the 41st Annual*. Waikoloa, HI: IEEE.

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. (2010). A social network system for analyzing publication activities of researchers. In *Symposium on collective intelligence, COLLIN 2010, Advances in Intelligent and Soft Computing*. Hagen, Germany: Springer.

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.

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

Bavelas, A. (1947). A mathematical model for group structures. *Human Organization*, *7*(3), 16–30.

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

Binder, J. J. (1985). On the use of the multivariate regression model in event studies. *Journal of Accounting Research*, *23*(1), 370–383.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*(1), 113–120.

Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social Networks*, *19*(3), 243–269.

Borgatti, S. P. (1995). Centrality and AIDS. *Connections*, *18*(1), 112–114.

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

Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, *69*(1), 169–173.

Burt, R. S. (1995). *Structural holes: The social structure of competition*. Belknap Pr.

Chung, K. S. K. (2009). *Understanding attitudes towards performance in knowledge-intensive work: The influence of social networks and ICT use*. University of Sydney.

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

Duque, R. B., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D. B. S., & 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.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239.

Freeman, L. C. (1980). The gatekeeper, pair-dependency and structural centrality. *Quality and Quantity*, *14*(4), 585–592.

Gibbons, M. R. (1982). Multivariate tests of financial models: A new approach. *Journal of Financial Economics*, *10*(1), 3–27.

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

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

Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly*, *44*(1), 82–85.

Hirsch, J. E. (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. H. (2006). *H*-Index: An evaluation indicator proposed by scientist. *Science Focus*, *1*(1), 8–9.

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

Kim, K., & Altmann, J. (2010). Measuring and analyzing the openness of the Web2.0 service network for improving the innovation capacity of the Web2.0 system through collective intelligence. In *Symposium on collective intelligence, COLLIN 2010, Advances in intelligent and soft computing*. Hagen, Germany: Springer.

Kousha, K., & Thelwall, M. (2007). Google Scholar citations and Google Web/URL citations: A multi-discipline exploratory analysis. *Journal of the American Society for Information Science and Technology*, *58*(7), 1055–1065.

Krackhardt, D. (1992). The strength of strong ties: The importance of philos in organizations. *Networks and Organizations: Structure, Form, and Action*, 216–239.

Leavitt, H. J. (1951). Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology*, *46*(1), 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. D., & Lautrup, B. E. (2006). Measures for measures. *Nature*, *444*(7122), 1003–1004.

Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science*, 1477–1490.

Lu, Y., Luo, X., Polgar, M., & Cao, Y. (2010). Social network analysis of a criminal hackers community. *Journal of Computer Information Systems*, (Winter), 31–41.

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

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

Melin, G. (2000). Pragmatism and self-organization: Research collaboration on the individual level. *Research Policy*, *29*(1), 31–40.

Owen-Smith, J., Riccaboni, M., Pammolli, F., & Powell, W. W. (2002). A comparison of US and European university–industry relations in the life sciences. *Management Science*, *48*(1), 24–43.

Prathap, G. (2006). Hirsch-type indices for ranking institutions scientific research output. *Current Science (Bangalore)*, *91*(11), 1438.

Reagans, R., & Zuckerman, E. W. (2001). Networks diversity, and productivity: The social capital of corporate R&D teams. *Organization Science*, 502–517.

Ruane, F., & Tol, R. S. J. (2008). Rational (successive) *h*-indices: An application to economics in the Republic of Ireland. *Scientometrics*, *75*(2), 395–405.

Ruhnau, B. (2000). Eigenvector-centrality—A node-centrality? Colloboration 1. *Social Networks*, *22*(4), 357–365.

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, *31*(4), 581–603.

Schubert, A. (2007). Successive *h*-indices. *Scientometrics*, *70*(1), 201–205.

Scott, J. (1991). *Social network analysis: A handbook*. Sage.

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. H. (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.
Suresh, V., Raghupathy, N., Shekar, B., & Madhavan, C. E. (2007). Discovering mentorship information from author collaboration networks. *Discovery Science*, *4755*, 197–208.
Tol, R. S. J. (2008). A rational, successive *g*-index applied to economics departments in Ireland. *Journal of Informetrics*, *2*(2), 149–155.
US Office of Science & Technology Policy. (2000). *Examples of international scientific collaboration and the benefits to society*. Retrieved June 14, 2005 from http://clinton4.nara.gov/WH/EOP/OSTP/html/00426_7.html
van Raan, A. F. J. (2004). Measuring science. Capita select a of current main issues. In *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.