Technology Management, Economics and Policy Papers

SEOUL NATIONAL UNIVERSITY
College of Engineering
Technology Management, Economics and Policy Program

On the Correlation between Research Performance and Social Network Analysis Measures Applied to Research Collaboration Networks

Alireza Abbasi, Jörn Altmann

TEMEP Discussion Paper No. 2010:66

서 울 대 학 교
공 과 대 학, 기술경영경제정책 대학원과정
151-742 서울시 관악구 관악로 599

Technology Management, Economics, and Policy Program
College of Engineering, Seoul National University
599 Gwanak-Ro, Gwanak-Gu, Seoul 151-742, South-Korea
Phone: ++82-2-880-9140, Fax: ++82-2-880-8389

Technology Management, Economics and Policy Program (TEMEP) is a graduate program at Seoul National University. It includes both M.Sc. and Ph.D. programs, which are integrated into the graduate programs at the College of Engineering.

The *TEMEP Discussion Paper Series* is intended to serve as an outlet for publishing research about theoretical, methodological, and applied aspects of industrial economics, especially those related to the institute's areas of specialization, namely management of technology, information and communication technology, telecommunication, services, health industries, energy industries, as well as infrastructures for development. In particular, paper submissions are welcome, which analyze current technology and industry related issues, discuss their implications, and suggest possible alternative policies. The objective is to gain insights into important policy issues and to acquire a balanced viewpoint of policy making, technology management, and economics. It will enable us to identify the problems in industries accurately and to come up with optimal and effective guidelines. In addition, another important aim of this series is to facilitate communication with external research institutes, individual researchers, and policy makers worldwide.

Research results disseminated by TEMEP may include views on policies for information and communication technologies, technology management, telecommunication, energy, health, and development economics, but the institute itself takes no institutional policy position. If appropriate, the policy views of the institute are disseminated in separate policy briefs. Thus, any opinion expressed in the TEMEP Discussion Paper Series is those of the author(s) and not necessarily the opinion of the institute.

Finally, the current editor of this series would like to thank Prof. Dr. Almas Heshmati for establishing this working paper series in January 2009 and for guiding its development during the course of the first 55 issues. In particular, Prof. Heshmati provided invaluable contributions for setting up the local IT infrastructure, registering the series with RePEc, disseminating the working papers, and setting the quality requirements.

Jörn Altmann, Editor

Office: 37-305
Technology Management, Economics and Policy Program
College of Engineering
Seoul National University
599 Gwanak-Ro, Gwanak-Gu
Seoul 151-742
South-Korea

Phone: +82-70-76787-6676
Fax: +1-501-641-5384
E-mail: jorn.altmann@acm.org

On the Correlation between Research Performance and Social Network Analysis Measures Applied to Research Collaboration Networks

Alireza Abbasi, Jörn Altmann

Technology Management, Economics, and Policy Program (TEMEP)
Department of Industrial Engineering
College of Engineering, Seoul National University
599 Gwanak-Ro, Gwanak-Gu
Seoul 151-742, South-Korea
abbasi@snu.ac.kr, jorn.altmann@acm.org

October, 2010

Abstract: In this study, we develop a theoretical model based on social network theory to understand how the collaboration (co-authorship) network of scholars correlates to the research performance of scholars. For this analysis, we use social network analysis (SNA) measures (i.e., normalized closeness centrality, normalized betweenness centrality, efficiency, and two types of degree centrality). The analysis of data shows that the research performance of scholars is positively correlated with two SNA measures (i.e., weighted degree centrality and efficiency). In particular, scholars with strong ties (i.e., repeated co-authorships, i.e., high weighted degree centrality) show a better research performance than those with low ties (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 (i.e., co-authors that have joined publications), perform better than those researchers with many relationships to the same group of linked co-authors.

Keywords: Social Network Analysis, Co-authorship Network, Researchers' Performance.

JEL Classification Numbers: C43, C44, C65, D80, D85, M12, M21

1. Introduction

1.1. Performance

Performance appraisal is an inevitable function of management at any level. It fosters the development progress. Consequently, within a research environment, there should also be a performance evaluation for academics at universities and research institutes. This evaluation of researchers, which should be based on their output and productivity, 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. 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 for a specific project to a scientific research group, it is important to choose the most appropriate scholars with the aim of maximizing the research output, cost savings, and resource utilization. However, in all these cases, the same problem exists, namely answering the question of how can the most suitable scientists, who can achieve the goals, be identified [1].

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 [2]. 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) [3]. 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" [3]. In other words, a scholar with an index of h has published h papers, which have been cited by others at least h times. The h-Index is also being used by many academic databases (e.g., *Web of Science*[1] and *Scopus*[2]) to measure the performance of scholars. Furthermore, the h-Index became also the basis for a wide range of new measures [4, 5, 6, 7, 8, 9]. There are some studies that suggest measures for evaluating the output of research communities by extending the previously mentioned indices to groups [4, 7, 10, 11, 12].

1.2. Collaboration

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 et al. as well as Wanger and Leydesdorff [13, 14, 15]. By jointly publishing a paper, researchers show their knowledge sharing activities, which are essential for knowledge creation. "The rising awareness of collaborativeness in science has led to a sharpened focus on the collaboration issue" [16]. Scientific collaboration has even been called a "springboard for economic prosperity and sustainable development" [17]. As most scientific output is a result of

---

[1] science.thomsonreuters.com/training/wos/(Citation Report)

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

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 [18].

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 [19, 20]. The appropriate design of research policies remains a major issue though [21].

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 [22]. With respect to the number of new publications, empirical studies have been conducted by Lee and Bozeman as well as Duque et al. [23, 24]. Although Duque et al. have found that collaboration was not associated with an increase in scientific publications in the developing countries of Ghana, Kenya, and India (Kerala) [24], Lee and Bozeman show that the total number of publications for US scientists is positively associated with the total number of collaborations [23].

1.3. Social networks

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" [25], 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 collaboration in multiple ways [25].

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 [26, 27]. 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 [25, 28].

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 [27]. 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 [29].

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.

1.4. Methodology

Currently, it is not clear which collaboration data is useful for evaluating the academic community. Although there is a large set of potential collaboration data, which qualifies for being used as a measure (e.g., joined conference organization, joined research proposal submissions, joined publications, joined conference attendance, and teacher-student relationships), we only considered joined publications as a measure in our study.

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 (h-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 three research questions:

- Which social network analysis measures can be used to evaluate the co-authorship-based research collaboration network of researchers?
- Does a correlation between the social network analysis measures and the h-Index exist?
- How can researchers and research communities improve their productivity?

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 [27].

After preparing the social network matrix, we used UCINET [43] as a tool for visualizing the network and for calculating network measurements including degree centrality, betweenness centrality, closeness centrality, and efficiency of each node. For correlating these measures and the research performance measure (i.e., h-Index), we use UCINET. The correlation uses the Spearman correlation test.

Our paper shows the results for five co-authorship networks, representing the collaborations of researchers of each of the five iSchools.

1.5. Paper organization

In the following chapter, based on a literature review about the connection between social network analysis measures and performance of actors, we introduce our research questions and our model for correlating SNA measures and research performance. Chapter 3 describes the data resources and the method of data gathering and validation. Chapter 4 shows the collaboration network of one of the ischools as an example. Besides, it presents the results of the social network analysis measures and the performance measure. Finally, we discuss the results, the research limitations, and our future work.

2. Model for correlating SNA measures and research performance

2.1. Co-authorship network

The co-authorship network, i.e., the research collaboration network, is represented through a graph as shown in the example of Figure 1.

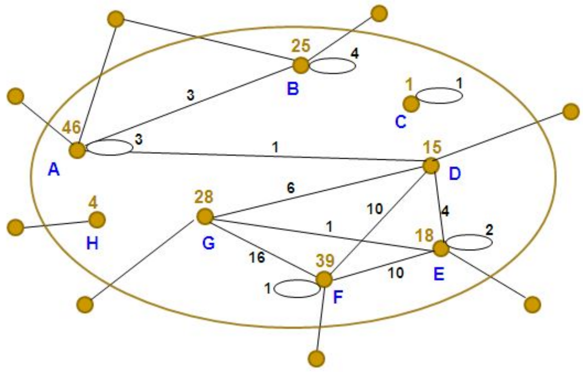

Figure 1. An example of co-authorship network of an academic community

The nodes (actors, participants, vertices) i of the graph represent researchers. The node weights w_i denote the total number of publications by a researcher. Links (ties, relations, edges) a_{ij} between node i and node j indicate collaboration relationships between nodes and represent the co-authorships 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 links w_{ij} denotes the number of publications that two researchers co-authored.

2.2. Centrality measures

A method used to understand the value, importance, and influence of actors in the networks is to evaluate the centrality of actors in the network.

Freeman reviewed and unified various measures of centrality [30]. In particular, Freeman defined centrality in terms of node degree centrality, betweenness centrality, and closeness, each having important implications on outcomes and processes [31]. For example, Freeman found that centrality is an important structural factor that influences leadership, satisfaction, and efficiency. Based on his study, it has been shown that betweenness centrality and degree centrality influence the performance of the actor.

Degree centrality is an indicator of an actor's communication activity [32]. The normalized degree centrality is defined as the number of links of an actor divided by the maximal possible number. The normalized degree centrality d_i of node i is given as:

$$d_i = \frac{\sum_j a_{ij}}{(n-1)} \quad ,$$

where a_{ij} indicates the existence or none-existence of a link between node i and node j. n represents the number of nodes. If there is any link between node i and node j, $a_{ij} = 1$. If there is no link, $a_{ij} = 0$.

Closeness centrality indicates the extent to which an actor is close to all others in the network [32]. It is defined as the inverse of the total graph-theoretic distance of a given node from all other nodes [33]. More precisely, the normalized closeness centrality c_i of node i is defined as:

$$c_i = \frac{(n-1)}{\sum_j e_{ij}} \quad ,$$

where n is the number of nodes and e_{ij} is the number of links in the shortest path from node i to node j. Closeness is an inverse measure of centrality in that a large value indicates a less central node, while a small value indicates a central node. It is a measure for the cost of communicating with other nodes in the network.

Betweenness is an indicator of an actor's potential control of communication within the network. Betweenness centrality is defined as the ratio of the number of shortest paths (between all pairs of nodes) that pass through a given node divided by the total number of shortest paths. The normalized betweenness centrality b_i of node i is given as:

$$b_i = \sum_{j,k \wedge i \neq j \neq k} \frac{g_{jik}}{g_{jk}} \Big/ \frac{(n-1)(n-2)}{2} \quad ,$$

where n is the number of nodes, g_{jk} is the number of shortest paths from node j to node k, and g_{jik} is the number of shortest paths from node j to node k that pass through node i.

In line with these definitions, it is expected that authors, which have a high degree centrality, a high betweenness centrality, and a low closeness value, are have a high potential of a good research performance as they are in the center of network. Since we measure research performance using the h-Index [3], as explained earlier, we can formulate these expectations as the following three hypotheses:

(H1a): *Normalized degree centrality of a researcher positively correlates to her h-Index.*

(H1b): *Normalized closeness centrality of a researcher positively correlates to her h-Index.*

(H1c): *Normalized betweenness centrality of a researcher positively correlates to her h-Index.*

2.3. Degree centrality for weighted graphs

Another way to analyze actors of a network has been introduced by Granovetter [34]. He established the theory of the 'Strength of Ties'. Besides, he argued that individuals obtain 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 a bridge to different clusters of people [32]. Granovetter 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" [37]. Krackhardt showed that strong ties are also important, especially in the generation of trust [35]. In addition to this, Levin and Cross [36] 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, controlling 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 [37, 38]. This work has been extended by analyzing the different effects of groups of actors on knowledge sharing [44].

Because of this theory, it is necessary to investigate whether the strength of a tie impacts the research performance. Since links of our research collaboration network are weighted (i.e., they represent the number of co-authorships between two scholars), we define the strength of a tie (link) between node i and node j as the weight of the link w_{ij} between those nodes.

Based on this definition, we can factor in the tie strength into the degree centrality of a node. Therefore, we calculate the average link weights of an actor's co-authorships (links). That means, we divide the sum of a node's link weights (number of co-authorships) by the total number of different co-authors. The degree centrality for the weighted graph (weighted degree centrality) d'_i is expressed as follows:

$$d'_i = \sum_j \frac{w_{ij}}{d_i(n-1)} \quad ,$$

where w_{ij} represents the weights of the links between node i and node j. $d_i\ (n-1)$ represents degree centrality.

Thus, based on these arguments, it is expected that scholars with a strong relationship (frequent co-authorship) achieve a high research performance. This is formulated as the following hypothesis:

(H2): *Weighted degree centrality of a researcher positively correlates to her h-Index.*

2.4. Structural holes

Holes in the network refer to the absence of ties (links) that would otherwise connect unconnected clusters together. Individuals, who bridge these holes attain an advantageous position that yields information and control benefits" [39]. Structural holes theory is based on betweenness centrality. The theory states that power and influence accrue to those actors, who broker connections between unconnected groups of people [32].

With respect to network optimization, Burt claims that "increasing network size (number of direct contacts) without considering the diversity reached by the contacts makes the network inefficient in many ways" [40]. Burt suggests taking advantage of structural holes when a network is planned to be increased. The absence of links between a node and its non-redundant nodes makes a structural hole.

Therefore, the number of non-redundant contacts (e.g., the four nodes of network B that are connected to the center node within Figure 2) is important to the extent that non-redundant contacts lead to people that could provide non-redundant information. In general, the idea is that “actors are in a better position to benefit from interactions with others, if they are connected to others, who are not well-connected themselves or are not well-organized” [32]. The reason is that “the extent of information coming from closely knit clusters tends to become redundant,” which makes networks inefficient [39].

Therefore, Burt’s efficiency concerns the number of groups of primary contacts (i.e., directly connected nodes) that are not connected to any other groups of primary contacts [39]. The network A in Figure 2 is inefficient as the center node (labeled “YOU”) gets from its 12 primary contacts 75% redundant information. Redundant information is spread by nodes belonging to the same cluster. Therefore, it can be stated that this node wastes its resources by maintaining its ties to all actors of the same cluster. Non-redundant collaborators, however, give access to diversity of information, which usually leads to innovation and high performance.

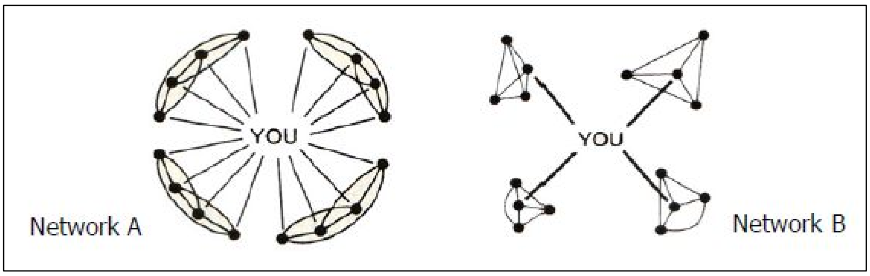

Figure 2. Two networks with structural holes, where Network A is less efficient than Network B (adapted from Chung [40] and Burt [39])

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 [32, 39]. Here, efficiency is defined as the ratio of the total number of disjunct groups of primary nodes, 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 primary nodes number (node degree).

$$f_i = \frac{g_i}{\sum_j a_{ij}} \quad ,$$

where g_i denotes the number of disjunct groups of primary contacts.

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. Therefore, in order to test this property, we formulated the following hypothesis:

(H3): *Efficiency of a researcher positively correlates to her h-Index.*

2.5. Model

Based on the results of the previous sections, we propose the following model (Figure 3). In order to capture a scholars’ collaboration activities, we calculate degree centrality, betweenness centrality, closeness centrality, tie strengths, and efficiency, based on the co-authorship network, i.e., the research collaboration network.

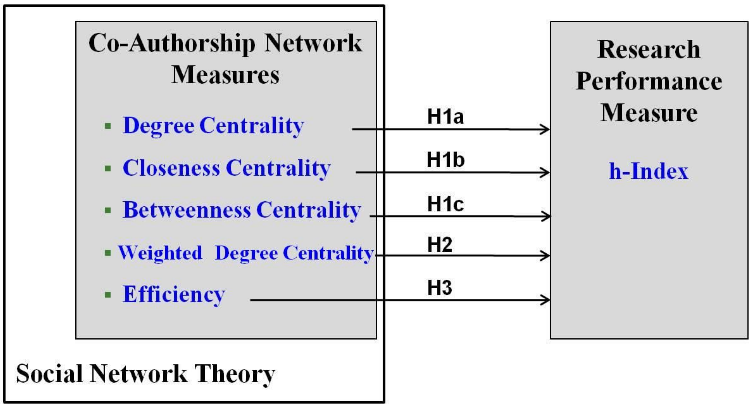

Figure 3. Research model to investigate effect of network measures on performance

3. Data collection

For this study, we collected data on five information schools (iSchools): University of Pittsburgh, UC Berkeley, 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, because of the fact, that 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 researchers, 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 [27]. AcaSoNet is a Web-based application for extracting publication information (i.e., 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) [41]. 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 show that rankings based on Google Scholar have a high rank correlation with rankings based on Web of Science or Scopus [42].

For our analysis, we followed Google Scholars approach and did not differentiate between the different types of publications (i.e., proceedings of local conferences, proceedings of international conferences, journals, books, and presentations were weighted equally). Our data covered a period of five years (2001 to 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.

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 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, 2139 publications, 1815 authors, and 5310 co-authorships were finally available for our analysis. For the analysis of the collaboration network of each ischool, we only considered professors and lecturers of each iSchool. In total, there were 132 professors and lecturers.

4. Analysis and results

Based on the available publication data of researchers, we can build a network matrix for each of the iSchools. Figure 4 shows the network matrix for the University of Pittsburgh iSchool. These matrices are the basis for our social network analysis. After importing one of the network matrices as a table into UCINET, we can start the analysis.

	Alma	Biagi	Brusi	Carb	Cox	Detle	He	Kaba	Kari	Krish	Lars	Lewi	Sprin	Tho	Tippe	Tom	Weis	Zado
S. Alman	8	1	-	1	1	1	1	2	-	-	-	-	-	1	-	2	1	-
M. Biagini	1	2	-	1	1	1	-	-	-	-	-	-	-	1	-	1	1	-
Brusilovsk	-	-	52	-	-	-	-	-	-	-	-	1	1	-	-	-	-	-
T. Carbo	1	1	-	12	1	1	-	-	-	-	-	-	-	1	-	1	2	-
R. Cox	1	1	-	1	89	1	-	-	-	-	-	-	-	1	-	1	1	-
Detlefsen	1	1	-	1	1	15	-	-	-	-	-	-	-	1	-	1	1	-
D. He	1	-	-	-	-	-	19	-	-	-	-	-	-	-	-	-	-	-
J. Kabara	2	-	-	-	-	-	-	18	-	8	-	-	-	-	8	-	-	-
H. Karimi	-	-	-	-	-	-	-	-	16	1	-	-	-	-	-	-	-	-
Krishnam	-	-	-	-	-	-	-	8	1	29	-	-	-	-	6	-	-	2
R. Larsen	-	-	-	-	-	-	-	-	-	-	9	-	-	-	-	-	1	-
M. Lewis	-	-	1	-	-	-	-	-	-	-	-	42	-	-	-	-	-	-
M. Spring	-	-	1	-	-	-	-	-	-	-	-	-	11	-	-	-	-	-
Thompso	1	1	-	1	1	1	-	-	-	-	-	-	-	12	-	1	1	-
D. Tipper	-	-	-	-	-	-	-	8	-	6	-	-	-	-	25	-	-	-
C. Tomer	2	1	-	1	1	1	-	-	-	-	-	-	-	1	-	7	1	-
M. Weiss	1	1	-	2	1	1	-	-	-	-	1	-	-	1	-	1	20	-
Zadorozh	-	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	18

Figure 4. Co-authorship relationships of University of Pittsburgh iSchool professors

UCINET allows visualizing the matrices (Figure 5). For small networks, this feature helps analysts to identify visually certain characteristics (e.g., the location of people within the network, and the network structure) and, then, initiate the calculation of SNA measures to investigate the characteristics in detail.

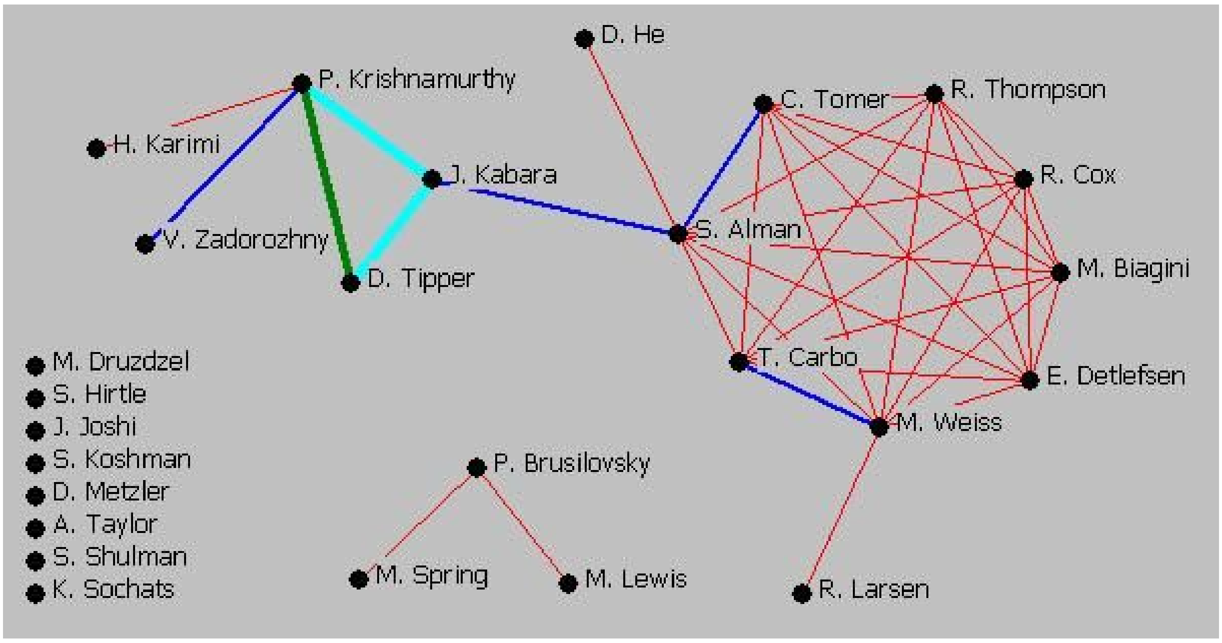

Figure 5. UCINET visualization of the co-authorship network of the University of Pittsburgh iSchool professors

Figure 5 demonstrates that the graph of the Pittsburg professors' co-authorship network is disconnected (i.e., there are ten sub-networks). The sub-networks are eight isolated nodes, a network of 3 nodes, and a network of 15 nodes. The eight isolated nodes represent authors that have no co-authorships with other scholars of the iSchool. With respect of the large sub-network, there is a core of eight people, which are strongly interconnected. The different widths of the links in Figure 5 represent different weights of links (i.e, number of co-authorships).

In addition, by looking at the network graph of Figure 5, we detect the strategic positions of 'S. Alman' and 'J. Kabara' within the network. The link between 'S. Alman' and 'J. Kabara' is a bridge. If the link were missing, a structural hole in the network would occur, i.e. the network would comprise another isolated sub-network. Individuals, who bridge these holes attain an advantageous position that yields information and control benefits [45]. 'J. Kabara' and 'S. Alman' have such a position. They bridge four actors, which are reachable by 'J. Kabara', and nine actors being reachable by 'S. Alman'.

Now, we calculate the SNA analysis measures of our model and the h-Index for all scholars of all five iSchools. The results for the University of Pittsburgh iSchool are shown in Table 1.

Table 1. Name, normalized degree centrality, normalized closeness centrality, normalized betweenness centrality, efficiency, weighted degree centrality, and h-Index of University of Pittsburgh iSchool researchers

	Name	N. Degree Centrality (%)	N. Closeness Centrality (%)	N. Betweenness Centrality (%)	Efficiency	Weighted Degree Centrality	h-Index
1	Susan W. Alman	33.33	7.01	15.1	0.33	1.08	3
2	Mary K. Biagini	25.93	6.91	0	0.14	1.00	1
3	Peter Brusilovsky	7.41	3.85	0.29	1.00	2.06	17
4	Toni Carbo	25.93	6.91	0	0.14	1.08	2
5	Richard J. Cox	25.93	6.91	0	0.14	1.05	9
6	Ellen Detlefsen	25.93	6.91	0	0.14	1.00	3
7	Marek Druzdzel	0	3.57	0	0.00	1.58	7
8	Daqing He	3.70	6.78	0	1.00	1.86	6
9	Stephen C. Hirtle	0	3.57	0	0.00	1.00	4
10	James B.D. Joshi	0	3.57	0	0.00	2.80	8
11	Joseph Kabara	11.11	6.92	11.40	0.67	2.00	4
12	Hassan Karimi	3.70	6.57	0	1.00	1.53	6
13	Sherry Koshman	0	3.57	0	0.00	3.00	2
14	Prashant Krishnamurthy	14.82	6.78	7.12	0.75	1.88	8
15	Ronald Larsen	3.70	6.70	0	1.00	1.36	3
16	Michael Lewis	3.70	3.84	0	1.00	3.31	13
17	Douglas Metzler	0	3.57	0	0.00	1.00	1
18	Stuart Shulman	0	3.57	0	0.00	2.92	6
19	Kenneth M. Sochats	0	3.57	0	0.00	1.75	1
20	Michael B. Spring	03.70	3.84	0	1.00	1.17	3
21	Arlene G. Taylor	0	3.57	0	0.00	1.00	3
22	Richard A. Thompson	25.93	6.91	0	0.14	1.00	2
23	David Tipper	07.41	6.75	0	0.50	2.26	6
24	Christinger Tomer	25.93	6.91	0	0.14	1.25	4
25	Martin B.H. Weiss	29.63	6.92	3.70	0.25	1.65	5
26	Vladimir Zadorozhny	0.0370	6.57	0	1.00	2.71	8

Based on the data of Table 1 and the SNA results for the other five iSchools, we test our hypotheses. For this, we use the Spearman correlation test. In particular, we calculate the Spearman correlation coefficient between all SNA measures and the performance index (h-Index), using UCINET. The results are shown in Table 2.

Table 2. Spearman correlation coefficients for professors of five iSchools

	h-Index	Normalized Degree Centrality	Normalized Closeness Centrality	Normalized Betweenness Centrality	Efficiency	Weighted Degree Centrality
h-Index	1					
Normalized Degree Centrality	0.093	1				
Normalized Closeness Centrality	0.064	0.490*	1			
Normalized Betweenness Centrality	0.117	0.424*	0.352*	1		
Efficiency	0.193*	0.209*	0.174*	0.112	1	
Weighted Degree Centrality	0.373*	0.053	0.137	0.097	0.202*	1

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

As the results in Table 2 show, the Spearman correlation coefficient does not show significance at the 0.05 level for all three normalized centrality measures (normalized degree centrality, normalized closeness centrality, and normalized betweenness centrality). Although literature suggested a correlation between these centrality measures and research performance, our results do not support it. Therefore, we cannot accept the first three hypotheses (H1a, H1b, and H1c).

Table 2 also shows that the weighted degree centrality has a positively significant correlation (cc = 0.373) with the performance measure (h-Index). A positively significant correlation (cc = 0.193) also exists between the efficiency measure and the h-Index. Therefore, we can accept the hypothesis H2 and the hypothesis H3. Consequently, we can state that having a high weighted degree centrality and a high efficiency represents a high research output.

Both positively significant correlations expose that researchers, who have strong ties (i.e., repeated co-authorships, i.e., high weighted degree centrality) 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 [35], which has been explained in chapter 2, has been supported by our analysis. Besides, the positive correlation between h-Index and efficiency also shows that researchers have to be selective about the ties that they maintain, following Burt [39]. In particular, a scholar should maintain non-redundant co-authorship relationships and, therefore, should focus on maintaining strong relationships to only one co-author of a group of linked co-authors (i.e., co-authors that have also joined publications).

Concluding, we can state that scholars should keep strong relationships with existing co-authors and build on former co-authorships. However, in order to increase the efficiency, they should only keep strong relationships with one of the co-authors of a group of linked co-authors.

5. Conclusion

In order to improve the benefit from research (and research funding), well-performing researchers have to be identified. As past research has shown, the h-Index can be a surrogate for evaluating the research performance of scholars [46]. In addition to this, the collaboration skills of researchers became more and more important over the past years.

However, it is still open 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. As social network analysis measures, the normalized degree centrality, the normalized closeness centrality, the normalized betweenness centrality, the weighted degree centrality, and the efficiency were considered.

The results of our analysis show that research performance is positively associated with weighted degree centrality and with efficiency. Scholars with strong ties (i.e., repeated co-authorships, i.e., high weighted degree centrality) 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.

Furthermore, access to demographic information of researchers (e.g., age, gender, and nationality) would 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.

In the future, we will extend our work by applying this methodology to groups of researchers in addition to individuals. For example, it would allow evaluating the research performance of entire departments within a university.

6. References

[1] Y. Jiang, "Locating active actors in the scientific collaboration communities based on interaction topology analyses," Scientometrics, vol. 74, no. 3, pp. 471-482, 2008.

[2] S. Lehmann, A. Jackson, and B. Lautrup, "Measures for measures," Nature, vol. 444, no. 7122, pp. 1003-1004, 2006.

[3] J. Hirsch, "An index to quantify an individual's scientific research output," Proceedings of the National Academy of Sciences, vol. 102, no. 46, pp. 16569, 2005.

[4] J. Altmann, A. Abbasi, and J. Hwang, “Evaluating the productivity of researchers and their communities: The RP-index and the CP-index,” International Journal of Computer Science and Applications, vol. 6, no. 2, pp. 104–118, 2009.

[5] L. Egghe, “Theory and practise of the g-index,” Scientometrics, vol. 69, no. 1, pp. 131-152, 2006.

[6] B. Jin, “H-index: an evaluation indicator proposed by scientist,” Science Focus, vol. 1, no. 1, pp. 8-9, 2006.

[7] R. Tol, “A rational, successive g-index applied to economics departments in Ireland,” Journal of Informetrics, vol. 2, no. 2, pp. 149-155, 2008.

[8] A. Sidiropoulos, D. Katsaros, and Y. Manolopoulos, “Generalized Hirsch h-index for disclosing latent facts in citation networks,” Scientometrics, vol. 72, no. 2, pp. 253-280, 2007.

[9] P. Batista, M. Campiteli, and O. Kinouchi, “Is it possible to compare researchers with different scientific interests?” Scientometrics, vol. 68, no. 1, pp. 179-189, 2006.

[10] G. Prathap, “Hirsch-type indices for ranking institutions' scientific research output,” CURRENT SCIENCE-BANGALORE-, vol. 91, no. 11, pp. 1438, 2006.

[11] T. Braun, W. Glänzel, and A. Schubert, “A Hirsch-type index for journals,” Scientometrics, vol. 69, no. 1, pp. 169-173, 2006.

[12] A. Schubert, “Successive h-indices,” Scientometrics, vol. 70, no. 1, pp. 201-205, 2007.

[13] T. Luukkonen, O. Persson, and G. Sivertsen, “Understanding patterns of international scientific collaboration,” Science, Technology & Human Values, vol. 17, no. 1, pp. 101, 1992.

[14] T. Luukkonen, R. Tijssen, O. Persson et al., “The measurement of international scientific collaboration,” Scientometrics, vol. 28, no. 1, pp. 15-36, 1993.

[15] C. Wagner, and L. Leydesdorff, “Network structure, self-organization, and the growth of international collaboration in science,” Research policy, vol. 34, no. 10, pp. 1608-1618, 2005.

[16] G. Melin, “Pragmatism and self-organization:: Research collaboration on the individual level,” Research policy, vol. 29, no. 1, pp. 31-40, 2000.

[17] US Office of Science & Technology Policy, 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, 2000.

[18] M. Leclerc, and J. Gagné, “International scientific cooperation: The continentalization of science,” Scientometrics, vol. 31, no. 3, pp. 261-292, 1994.

[19] A. van Raan, “Measuring science. Capita selecta 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, 2004.

[20] J. Katz, and B. Martin, “What is research collaboration?,” Research policy, vol. 26, no. 1, pp. 1-18, 1997.

[21] G. Melin, and O. Persson, “Studying research collaboration using co-authorships,” Scientometrics, vol. 36, no. 3, pp. 363-377, 1996.

[22] D. Stokols, R. Harvey, J. Gress et al., “In vivo studies of transdisciplinary scientific collaboration:: Lessons learned and implications for active living research,” American Journal of Preventive Medicine, vol. 28, no. 2, pp. 202-213, 2005.

[23] S. Lee, and B. Bozeman, “The impact of research collaboration on scientific productivity,” Social Studies of Science, vol. 35, no. 5, pp. 673, 2005.

[24] R. B. Duque, M. Ynalvez, R. Sooryamoorthy, P. Mbatia, D.-B. Dzorgbo, and W. Shrum, “Collaboration paradox: Scientific productivity, the Internet, and problems of research in developing areas,” Social Studies of Science, vol. 35, no. 5, pp. 755, 2005.

[25] D. Sonnenwald, “Scientific collaboration: a synthesis of challenges and strategies,” Annual Review of Information Science and Technology, vol. 41, pp. 643-681, 2007.

[26] K. Kim, J. Altmann, and J. Hwang, “Measuring and analyzing the openness of the Web2.0 service network for improving the innovation capacity of the Web2.0 system through collective intelligence,” Symposium on Collective Intelligence, COLLIN 2010, in: Advances in Intelligent and Soft Computing, Springer, Hagen, Germany, 2010.

[27] A. Abbasi, and J. Altmann, “A social network system for analyzing publication activities of researchers,” Symposium on Collective Intelligence, COLLIN 2010, in: Advances in Intelligent and Soft Computing, Springer, Hagen, Germany, 2010.

[28] J. Owen-Smith, M. Riccaboni, F. Pammolli, and W. W. Powell, “A comparison of US and European university-industry relations in the life sciences,” Management Science, vol. 48, no. 1, pp. 24-43, 2002.

[29] A. Abbasi, J. Altmann, and J. Hwang, “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, vol. 83, no. 1, pp. 1-13, 2010.

[30] L. C. Freeman, “Centrality in social networks conceptual clarification,” Social Networks, vol. 1, no. 3, pp. 215-239, 1979.

[31] K. Chung, L. Hossain, and J. Davis, "Social networks and ICT correlates to Individual Performance." Applications of Social Network Analysis, pp. 1-16, 2005.

[32] K. Chung, and L. Hossain, “Measuring performance of knowledge-intensive workgroups through social networks,” Project Management Journal, vol. 40, no. 2, pp. 34-58, 2009.

[33] S. P. Borgatti, “Centrality and AIDS,” Connections, vol. 18, no. 1, pp. 112-114, 1995.

[34] M. S. Granovetter, “The strength of weak ties,” American Journal of Sociology, vol. 78, no. 6, pp. 1360, 1973.

[35] D. Krackhardt, “The strength of strong ties: The importance of philos in organizations,” Networks and Organizations: Structure, Form, and Action, pp. 216-239, 1992.

[36] D. Z. Levin, and R. Cross, “The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer,” Management science, pp. 1477-1490, 2004.

[37] M. T. Hansen, “The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits,” Administrative Science Quarterly, vol. 44, no. 1, pp. 82-85, 1999.

[38] R. Reagans, and E. Zuckerman, “Networks, diversity, and productivity: The social capital of corporate R&D teams,” Organization Science, pp. 502-517, 2001.

[39] R. S. Burt, “Structural holes: The social structure of competition, ” Belknap Pr, 1995.

[40] K. Chung. "Understanding attitudes towards performance in knowledge-intensive work: The influence of social networks and ICT use," http://hdl.handle.net/2123/4018.

[41] K. Kousha, and M. Thelwall, “Google Scholar citations and Google Web/URL citations: a multi-discipline exploratory analysis,” Journal of the American Society for Information Science and Technology, vol. 58, no. 7, pp. 1055-1065, 2007.

[42] F. Ruane, and R. Tol, “Rational (successive) h-indices: An application to economics in the Republic of Ireland,” Scientometrics, vol. 75, no. 2, pp. 395-405, 2008.

[43] S. Borgatti, M. Everett, and L. Freeman, “Ucinet for windows: Software for social network analysis (version 6),” Harvard, MA: Analytic Technologies, 2002.

[44] K. Kim, J. Altmann, and J. Hwang, “An Analysis of the openness of the Web2.0 service network using two sets of indices for measuring the impact of service ownership,” Hawaii International Conference on Systems Science, HICSS44, Koloa, Hawaii, USA, January 2011.

[45] R. S. Burt, “Structural holes and good ideas 1,” American journal of sociology, vol. 110, no. 2, pp. 349-399, 2004.

[46] C. Borgman, and J. Furner, “Scholarly communication and bibliometrics,” Annual Review of Information Science and Technology, vol. 36, pp. 3-72, 2002.

TEMEP Discussion Papers

2009-20: Marcel Risch and Jörn Altmann, "Enabling Open Cloud Markets Through WS-Agreement Extensions"

2009-21: Almas Heshmati and Rachid El-Rhimaoui, "Effects of Ownership and Market Share on Performance of Mobile Operators in MENA Region"

2009-22: Donghyun Oh, Almas Heshmati and Hans Lööf, "Total Factor Productivity of Korean Manufacturing Industries: Comparison of Competing Models with Firm-Level Data"

2009-23: Almas Heshmati, Dan Johansson and Carl Magnus Bjuggren, "Effective Corporate Tax Rates and the Size Distribution of Firms"

2009-24: Seong-Sang Lee and Yeonbae Kim, "The Impact of Efficiency Parameters on Firms' Innovative Activities: Evidence from Korean Firm-Level Data"

2009-25: Junseok Hwang, Dongook Choi, Jongeun Oh and Yeonbae Kim, "Competition in the Korean Internet Portal Market: Network Effects, Profit, and Market Efficiency"

2009-26: Jae Young Choi, Yeonbae Kim, Yungman Jun and Yunhee Kim, "A Multivariate Probit Analysis of Korean Firms' Information System Adoption: An Empirical Analysis on the Determinants of the Adoption and Complementarity among the Information Systems"

2009-27: Flávio Lenz-Cesar and Almas Heshmati, "Determinants of Firms Cooperation in Innovation"

2009-28: Firas H. Al-Hammadany and Almas Heshmati, "Determinants of Internet Use in Iraq"

2009-29: Jeong-Dong Lee, "Industry Dynamics and Productivity Research"

2009-30: Nam Tran-Quy and Almas Heshmati, "Measurements and Determinants of Pay Inequality and its Impacts on Firms Performance in Vietnam"

2009-31: Nils Karlson, Marcus Box and Almas Heshmati, "Generality, State Neutrality and Unemployment in the OECD"

2009-32: Chang Seob Kim, Yoonmo Koo, Ie-jung Choi, Junhee Hong and Jongsu Lee, "Consumer Preferences for Automobile Energy Efficiency Grades"

2009-33: Jongsu Lee, Jae Young Choi and Youngsang Cho, "A Forecast Simulation Analysis of the Next-Generation DVD Market Based on Consumer Preference Data"

2009-34: Jungwoo Shin, Chang Seob Kim and Jongsu Lee, "Model for Studying Commodity Bundling with a Focus on Consumer Preference"

2009-35: Yuri Park, Hyunnam Kim and Jongsu Lee, "Empirical Analysis of Consumer Preferences for Mobile Phones and Mobile Contents in the Presence of Indirect Network Effects"

2009-36: Jörn Altmann and Juthasit Rohitratana, "Software Resource Management Considering the Interrelation between Explicit Cost, Energy Consumption, and Implicit Cost: A Decision Support Model for IT Managers"

2009-37: Marcel Risch, Ivona Brandic and Jörn Altmann, "Using SLA Mapping to Increase Market Liquidity"

2009-38: Amit K. Bhandari and Almas Heshmati, "Willingness to Pay for Biodiversity Conservation"

2010-39: Marcel Risch, Jörn Altmann, Li Guo, Alan Fleming and Costas Courcoubetis, "The GridEcon Platform: A Business Scenario Testbed for Commercial Cloud Services"

2010-40: Junseok Hwang, Jörn Altmann and Ashraf Bany Mohammed, "Determinants of Participation in Global Volunteer Grids: A Cross-Country Analysis"

2010-41: Ruzana Davoyan, Jörn Altmann and Wolfgang Effelsberg, "Intercarrier Compensation in Unilateral and Bilateral Arrangements"

2010-42: Ruzana Davoyan, Jörn Altmann and Wolfgang Effelsberg, "Exploring the Effect of Traffic Differentiation on Interconnection Cost Sharing"

2010-43: Ruzana Davoyan and Jörn Altmann, "Investigating the Role of a Transmission Initiator in Private Peering Arrangements"

2010-44: Ruzana Davoyan, Jörn Altmann and Wolfgang Effelsberg, "A New Bilateral Arrangement between Interconnected Providers"

2010-45: Marcel Risch and Jörn Altmann, "Capacity Planning in Economic Grid Markets"

2010-46: Dang Minh Quan and Jörn Altmann, "Grid Business Models for Brokers Executing SLA-Based Workflows"

2010-47: Dang Minh Quan, Jörn Altmann and Laurence T. Yang, "Error Recovery for SLA-Based Workflows within the Business Grid"

2010-48: Jörn Altmann, Alireza Abbasi and Junseok Hwang, "Evaluating the Productivity of Researchers and their Communities: The RP-Index and the CP-Index"

2010-49: Jörn Altmann and Zelalem Berhanu Bedane, "A P2P File Sharing Network Topology Formation Algorithm Based on Social Network Information"

2010-50: Tai-Yoo Kim, Seunghyun Kim and Jongsu Lee, "The Gene of an Accelerating Industrial Society: Expansive Reproduction"

2010-51: Almas Heshmati and Sangchoon Lee, "The Relationship between Globalization, Economic Growth and Income Inequality"

2010-52: Flávio Lenz-Cesar and Almas Heshmati, "Agent-based Simulation of Cooperative Innovation"

2010-53: Erkhemchimeg Byambasuren and Almas Heshmati, "Economic Development in Mongolia"

2010-54: Almas Heshmati and Subal C. Kumbhakar, "Technical Change and Total Factor Productivity Growth: The Case of Chinese Provinces"

2010-55: Bory Seng and Almas Heshmati, "Digital Divide and Its Variations Amongst OECD, NIE and ASEAN Countries"

2010-56: Kibae Kim, Jörn Altmann and Junseok Hwang, "The Impact of the Subgroup Structure on the Evolution of Networks: An Economic Model of Network Evolution"

2010-57: Kibae Kim, Jörn Altmann and Junseok Hwang, "Measuring and Analyzing the Openness of the Web2.0 Service Network for Improving the Innovation Capacity of the Web2.0 System through Collective Intelligence"

2010-58: Alireza Abbasi and Jörn Altmann, "A Social Network System for Analyzing Publication Activities of Researchers"

2010-59: Jörn Altmann, Costas Courcoubetis and Marcel Risch, "A Marketplace and its Market Mechanism for Trading Commoditized Computing Resources"

2010-60: Fatmawati Zifa, Jörn Altmann, "A empirically validated framework for limiting free-riding in P2P network through the use of social network"

2010-61: Ashraf Bany Mohammed, Jörn Altmann, Junseok Hwang, "Cloud Computing Value Chains-Understanding Business and Value Creation in the Cloud"

2010-62: Ashraf Bany Mohammed, Jörn Altmann, "A Funding and Governing Model for Achieving Sustainable Growth of Computing e-Infrastructures"

2010-63: Junseok Hwang, Jihyoun Park, Jörn Altmann, "Two Risk-aware Resource Brokering Strategies in Grid Computing: Broker-driven vs. User-driven Methods"

2010-64: Juthasit Rohitratana, Jörn Altmann, "Agent-Based Simulations of the Software Market under Different Pricing Schemes for Software-as-a-Service and Perpetual Software"

2010-65: Radoslaw R. Okulski, Almas Heshmati, "Time Series Analysis of Global Airline Passengers Transportation Industry"

2010-66: Alireza Abbasi, Jörn Altmann, "On the Correlation between Research Performance and Social Network Analysis Measures Applied to Research Collaboration Networks"