

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

Alireza Abbasi · Jörn Altmann · Junseok Hwang

Received: 16 July 2008 / Published online: 31 December 2009
© Akadémiai Kiadó, Budapest, Hungary 2009

Abstract Although there are many studies for quantifying the academic performance of researchers, such as measuring the scientific performance based on the number of publications, there are no studies about quantifying the collaboration activities of researchers. This study addresses this shortcoming. Based on three measures, namely the collaboration network structure of researchers, the number of collaborations with other researchers, and the productivity index of co-authors, two new indices, the *RC-Index* and *CC-Index*, are proposed for quantifying the collaboration activities of researchers and scientific communities. After applying these indices on a data set generated from publication lists of five schools of information systems, this study concludes with a discussion of the shortcomings and advantages of these indices.

Keywords Collaborative networks · Collaboration activities · Social network analysis · Collaboration evaluation · Individual and community productivity · Collaboration measures · Indices · Empirical data analysis

JEL Classification C43 · D80 · D85 · L25 · M12 · M21

Introduction

The evaluation of research collaboration has obtained more attention during the past years for two major reasons. First, scholars realized that either local collaboration or collaboration over large distances via the Internet increases the research output. Many collaboration

A. Abbasi · J. Altmann (✉) · J. Hwang
Technology Management, Economics and Policy Program, Department of Industrial Engineering,
College of Engineering, Seoul National University, 599 Gwanak-Ro, Gwanak-Gu,
Seoul 151-742, South Korea
e-mail: jorn.altmann@acm.org

A. Abbasi
Research and Development Division, International Affairs,
National Iranian Oil Company, Tehran, Iran

activities result in publications of joined work, dissemination of research results, and the creation of new ideas (Borgman and Furner 2002). Melin shows that collaboration generates more knowledge, which, in turn, leads to new ideas (innovation) and academic productivity (Melin 2000). In his study, 38% of the academic respondents mentioned “Increased Knowledge” as the main benefit generated from joined publications (i.e. collaboration). It has also been observed that most scientific output is a result of group work and that most research projects are too large for an individual researcher to conduct. Consequently, scientific cooperation along with international scientific cooperation is needed more and more (Leclerc and Gagn 1994). Therefore, the increase in demand of researchers for an effective collaboration comes from the mutual benefit they gain.

Second, it is believed that collaboration leads to a decrease in redundancy of research efforts and increased resource savings (Jiang 2008), reducing the costs for research. Government policies exist that force researchers to collaborate, increase communication, and share their facilities (Ziman 1994). In general, it has been observed that research collaborations have become a major issue in science policy (Melin and Persson 1996). Collaboration as a policy objective became essential for countries to keep up with the scientific progress (Van Raan 2006).

Consequently, the evaluation and comparison of researchers based on their productivity became important. For example, the evaluation of researchers is being applied for allocating government and industry funding, recruiting faculty, ranking of universities, and for promoting faculty. Therefore, the objective of a researcher is to achieve a high evaluation by publishing many papers and getting a high citation count for publications (qualifying the quantity of publications) (Lehmann et al. 2006). Hirsch’s *h-Index* (2005) and Egghe’s *g-Index* (2006) are the most known measures for evaluating individual’s research performance. Subsequent research suggested many improvements of these two indices (Batista et al. 2006; Sidiropoulos et al. 2007; Jin 2006; Tol 2008; Altmann et al. 2009). In addition to this, there are some studies investigating the output of research communities. These studies extended these indices so that they can be applied to groups of researchers (Prathap 2006; Schubert 2007; Braun et al. 2005; Tol 2008; Altmann et al. 2009).

To analyze the collaboration activities of researchers, not only the research output of researchers (i.e. productivity) but also the collaboration network of researchers has to be measured. However, only a few studies on the social relationships between researchers, based on the scholarly publications of the researchers, have been published so far. Only two studies exist on social networks of research institutions. Among these two studies measuring the scholarly collaboration activities, there is only one that measures the number of interactions between researchers within a community (Jiang 2008). The other study considered the co-authorship between researchers within a community as the most important measure for collaboration (Laudel 2002). However, there are no measures that are “sensitive enough to reveal the structure and change of collaborative networks” (Melin and Persson 1996). In addition to this, it is also not clear which data is useful for evaluating academic communities. In particular, there is no index for a community which measures the co-authorship activities of scholars of the community. Tools and measures are missing as well.

Our study proposes two indices, one index for quantifying the collaboration activities of researchers (RC-Index) and one for community collaboration activities (CC-Index). These indices take into consideration the quantity of research output (i.e. number of different co-authors, and the number of publications with each co-author) and the quality of research output by using a researcher productivity index such as the RP-Index (Altmann et al. 2009).

There is little data available for evaluating the efficiency of research collaborations. To obtain data for our study, we used our Web-based tool, which is described in (Abbasi and

Altmann 2009). This tool extracts automatically publication lists of scholars from the Web and, then, allows scholars to modify and update this information. In a second step, the research collaboration network of users is created based on the collected data about publications and co-authorships.

The remainder of the paper is organized as follows. The next section describes the data collection procedure in detail. Section “[Methodology](#)” explains the methodology followed by the definition of the indices in section “[Evaluation of researchers](#)”. In section “[Results](#)”, we present our results. The final section concludes the paper with a discussion of the results and an outlook of future work.

Data collection

For this study, we collected data about five information management schools. These schools belong to the University of Pittsburgh (PITT), UC Berkeley (UCB), University of Maryland (UMD), University of Michigan (UMICH), and Syracuse University (SYR). These schools have been chosen, since they offer similar, relatively new programs in the area of information management and systems. The data sources, that contain the publication lists of researchers, are the annual reports of those schools, the *ACM portal*¹, *Google Scholar*², and *DBLP*³. Citation data has been collected from *Google Scholar* and the *ACM Portal*. For collecting, storing, and analyzing this data, we used AcaSoNet (Abbasi and Altmann 2009). AcaSoNet is a Web application for extracting publication information (e.g. author names, title, date of publication, publisher, and number of citations) from the Web, for identifying relationships between researchers (co-authorship of researchers), and for storing this data in a database.

Google Scholar considers a large variety of publication types (i.e. proceedings of local and international conferences, journals, books, and presentations), which can be found in freely accessible databases. Thus, it finds the most publications and citations per researcher compared to other indexing services (Kousha and Thelwall 2007). Consequently, the h-Index and g-Index calculated on Google Scholar data is higher than those run on data of *Web-of-Science* and *Scopus*. Nevertheless, Ruane and Tol (2008) show that rankings based on Google Scholar have a high rank correlation with rankings based on *Web of Science* or *Scopus*.

Following Google Scholar’s approach, we also did not differentiate between the different types of publications (i.e. all publications except presentations were treated equally). Our data covered 5 years, 2001–2005, except for the Maryland iSchool, for which we could not find data for the year 2002. We filled this vacancy with data of the year 2006. In total, 2,139 publications and 1,815 authors were extracted from the publication list of those five iSchools, showing 5,310 co-authorships.

Methodology

While there are many types of collaborations among researchers (e.g. co-authorship, co-partners in projects, and co-cited publications) (Jiang 2008), co-authorship can be

¹ <http://portal.acm.org/portal.cfm>.

² <http://scholar.google.com/>.

³ <http://www.informatik.uni-trier.de/~ley/db/>.

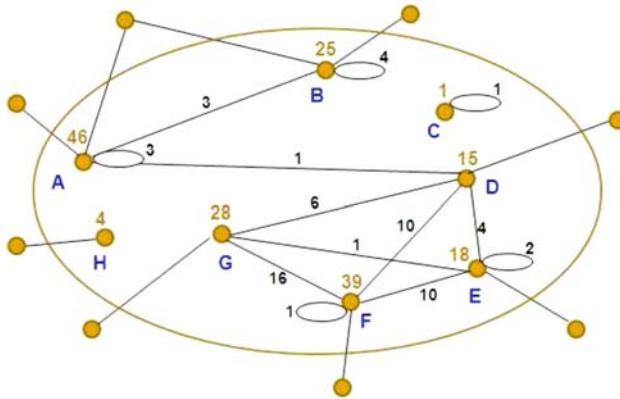


Fig. 1 Example for a collaboration network of an academic community

considered to be the most visible and accessible indicator for collaboration among scholars. Therefore, the co-authorship (especially in bibliometric studies) has often been used as a surrogate for collaboration (Borgman and Furner 2002). Within this study, they authors used the number of co-authorships and the number of different co-authors as indicators for collaborations between individual researchers as well as communities of researchers.

This study extends this notion. It proposes two new indices, namely the *Researcher Collaboration Index (RC-Index)* and the *Community Collaboration Index (CC-Index)*, which consider the quantity and quality of co-authorships as an indication of how powerful the collaboration of a researcher (respectively research community) is. For these indices, we use the *Researcher Productivity Index (RP-Index)*, indicating the amount of qualified publications by a researcher, and the *Community Productivity Index (CP-Index)*, indicating the amount of productive and qualified researchers in a community (The detailed descriptions of the RP-Index and CP-Index, that have been taken from Altmann et al. (2009), is given in Appendix 1).

For the definition of the two collaboration indices, we define a graph that represents the collaboration network structure of a group of researchers. An example of a graph is shown in Fig. 1. The nodes of the graph represent researchers. The node weights denote the number of publications published by a researcher. Links indicate co-authorships of authors of papers. Publications of which an author is the sole author are represented as loops in the graph. The weight of a link denotes the number of publications that the linked researchers co-authored.

By following the definition of this type of graph, the data set collected is preprocessed. As it can be seen in the remainder of this paper, the resulting graph helps us to answer the following two questions: First, who disseminates knowledge within a research community the most? Second, who has the strongest research network?

Evaluation of researchers

As a basic measure for collaboration activities of researchers, previous research used the *number of different co-authors* and the *number of collaborations* (i.e. the sum of link

weights) (Borgman and Furner 2002). We chose neither of these measures since they cannot reflect the quality of the collaboration. Even if we would account for the quality of a co-author by excluding co-authors with a low productivity (i.e. excluding co-authors, whose number of publications, number of citations, h-Index, or RP-Index is lower than a given threshold), setting the threshold value becomes an issue. Since this threshold value is an arbitrary number, it is difficult to argue for its validity.

For capturing the quality of collaboration activity, we introduce the *Co-author Collaboration Value (CCV)*. The CCV_a of author j and his co-author a is defined as the total number of collaborations between them multiplied by the co-author a 's RP-Index (which covers the quality of the co-author's research). The CCV_a indicates how productive the collaboration between co-author a and author j is. The definition of CCV_a is given by Eq. 1:

$$CCV_a = TotalNumberOfCollaborationsBetweenAuthors(j, a) * RP'_a. \quad (1)$$

Note, although we used RP'_a , the co-author collaboration value also works for the definition of the researcher productivity index (RP-Index) that gives natural numbers (RP_j). Further details about the definition of the RP-Index can be found in Appendix 1 (Altmann et al. 2009).

Based on the CCV definition, we can define an index, which quantifies the amount of collaborations with qualified (productive) co-authors. Given a list co-authors ranked in decreasing order of their CCV_a , the *Researcher Collaboration Index (RC-Index)* is defined as the largest natural number x , such that the top x co-authors of a researcher j have at least in average a value of x for their CCV_a . The RC-Index of researcher j can be written formally as shown in Eq. 2:

$$RC_j = \max(x \mid \frac{1}{x} \sum_{a=1}^x CCV_a \geq x), \quad x \in \mathbb{N}. \quad (2)$$

The definition of the RC-Index is a combination of principles used in the h-Index and g-Index. We preferred the RC-Index definition to potential other definitions that follow closer the h-Index or the g-Index definitions (Hirsch 2005; Egghe 2006).

Since the RC-Index generates natural numbers, it is likely that many researchers get the same RC-Index value. However, if we allow real numbers to be generated by the RC-Index, we can obtain a fine-grained evaluation of the performance of collaboration activities. Since real numbers are used in many different contexts to evaluate people (e.g. school, college, graduate school), it is also appropriate to evaluate researchers with real numbers. Following this line of argument, we obtain Eq. 3 by slightly modifying Eq. 2:

$$RC'_j = \begin{cases} \frac{1}{RC_j} \sum_{i=1}^{RC_j} CCV_a, & \text{if } \frac{1}{RC_j} \sum_{i=1}^{RC_j} CCV_a < RC_j + 1 \\ RC_j + 0.99, & \text{if } \frac{1}{RC_j} \sum_{i=1}^{RC_j} CCV_a \geq RC_j + 1 \end{cases}. \quad (3)$$

Equation 3 guarantees that RC'_j fulfills the condition $RC_j \leq RC'_j \leq RC_j + 1$ even if the researcher has very few co-authors with high CCV_a values. Within the remainder of the paper, in order to show the workings of these indices, we calculate the RC-Index giving natural numbers (Eq. 2) and real numbers (Eq. 3).

Evaluation of communities

A community can be any group of individuals. In the research context, an individual (i.e. researcher) can belong to different communities. For example, at a university, a researcher can belong to, in hierarchical order, a research group, a department, a program, a school, and a college. The university itself can even be considered a community. The performance of communities of each hierarchy level could be compared.

Based on the same concept that has been used for evaluating the collaboration activities of researchers, the *Community Collaboration Index (CC-Index)*, which measures the collaboration activities of a community, can be defined. The CC-Index is defined as follows. Given a set of researchers, which belong to a community and are ranked in decreasing order of their RC-Index, the CC-Index is the largest number such that the top x researchers have at least in average a value of x for their RC-Index. This index indicates the number of top researchers with a productive collaboration network.

$$CC_j = \max\left(x \mid \frac{1}{x} \sum_{a=1}^x RC_a \geq x\right), \quad x \in \mathbb{N}. \quad (4)$$

In the same way as for the researcher collaboration index, we can generate a more fine-grained evaluation of the community collaboration activity by using real numbers. Using the results of Eq. 4, we obtain the following equation for CC'_j :

$$CC'_j = \begin{cases} \frac{1}{CC_j} \sum_{a=1}^{CC_j} RC_a, & \text{if } \frac{1}{CC_j} \sum_{a=1}^{CC_j} RC_a < CC_j + 1 \\ CC_j + 0.99, & \text{if } \frac{1}{CC_j} \sum_{a=1}^{CC_j} RC_a \geq CC_j + 1 \end{cases}. \quad (5)$$

For our analysis of the different research communities (i.e. iSchools), we consider only Eq. 4.

Results

By applying the newly defined measures for assessing the collaboration activities of researchers and communities to our data set about the five iSchools, we can demonstrate the usefulness of our measures.

Evaluation results of the collaboration activity of iSchool researchers

For demonstrating the workings of the RC-Index, Table 1 shows the top 20 most collaborative researchers among all five iSchools. They are ranked in decreasing order of their RC-Index. The most collaborative researchers are from the *University of Michigan* (65%), which are followed by researchers from the *University of Maryland* (20%), *Syracuse University* (5%), *University of Pittsburgh* (5%), and *UC Berkeley* (5%).

In order to show how the RC-Index is calculated, we look at the data of one researcher in detail. We pick Elliot Soloway, who is ranked highest among all the scholars in our data set (Table 1). Elliot Soloway has a total of 34 co-authors and 104 collaborations during the observation period. For each of his co-authors a , we calculate the Co-author Collaboration Value (CCV_a), using Eq. 1. Although the calculation of the CCV_a of all of the co-authors is

Table 1 The top 20 most collaborative researchers of five iSchools^a

	Name	School ^a	RP'	Co-Au-No	Col-No	RC	RC'
1	Elliot Soloway	UMICH	6.99	34	104	18	18.02
2	Kevin Crowston	SYR	11.15	28	75	16	16.99
3	Joseph Krajcik	UMICH	4.99	29	70	15	15.99
4	Chris Quintana	UMICH	3.99	20	59	15	15.77
5	Mark W. Newman	UMICH	4.8	24	71	14	14.72
6	Jimmy Lin	UMD	5.57	49	108	13	13.82
7	Marti Hearst	UCB	7.99	23	58	13	13.03
8	W. Keith Edwards	UMICH	3.99	17	52	12	12.99
9	Judith S. Olson	UMICH	5.99	28	60	12	12.54
10	Jana Z. Sedivy	UMICH	2.99	12	42	12	12.45
11	Kathleen Luchini	UMICH	3.2	13	37	12	12.04
12	Steven P. Abney	UMICH	5.56	70	97	11	11.99
13	Michael Lewis	PITT	2.61	36	119	11	11.99
14	Benjamin B. Bederson	UMD	1.37	32	72	11	11.99
15	Gary M. Olson	UMICH	5.99	26	48	11	11.99
16	Cathleen Norris	UMICH	4.9	14	30	11	11.99
17	Martha E. Pollack	UMICH	9.99	39	70	11	11.9
18	Boris Katz	UMD	3.99	18	51	11	11.78
19	Dina Demner-Fushman	UMD	4.33	24	48	11	11.2
20	Nathan Bos	UMICH	2.99	13	35	11	11.17

The table shows the name of the researcher, the affiliation of the researcher (School), the researcher productivity index (RP'), the number of different co-authors (Co-Au-No), the total number of collaborations (Col-No), the Researcher Collaboration Index as a natural number (RC) and as a real number (RC')

^a iSchool of University of Pittsburgh (PITT), iSchool of UC Berkeley (UCB), iSchool of University of Maryland (UMD), iSchool of University of Michigan (UMICH), and iSchool of Syracuse University (SYR)

needed for calculating the RC-Index, Table 2 shows only the 20 co-authors with the highest CCV_a. Based on these results and Eq. 2, Elliot Soloway's RC-Index can be calculated to be 18.

To calculate the RC-Index of each researcher of each iSchool, similar data as shown in Table 2 (i.e. the CCV_a and the RP-Index of each co-author *a*) is needed for each researcher of each iSchool.

Table 3 shows the researcher collaboration index as a natural number (RC) and as a real number (RC') for the top 10 researchers of each iSchool. The researchers of each iSchool are ranked in decreasing order of their collaboration index (RC-Index). The RP-Index (RP') is also calculated, however, simply as a reference.

Evaluation results of the collaboration activity of communities

Table 4 shows the results for the community collaboration index (CC-Index), using the iSchool data set and the iSchool researcher collaboration index table (Table 3). Within Table 4, the last four columns show the community productivity index (CP and CP') and the community collaboration index (CC and CC'). Although the ranking according to the productivity index shows that the iSchool of Berkeley, Michigan, and Syracuse are the

Table 2 The top 20 co-authors of Elliot Soloway in decreasing order of their CCV_a

	Name of co-author	Number of collaborations	RP'	CCV_a	
	1	Joseph Krajcik	15	4.99	74.85
	2	Chris Quintana	18	3.99	71.82
	3	Cathleen Norris	11	4.90	53.90
	4	Kathleen Luchini	11	3.20	35.20
	5	Terry Sullivan	4	3.02	12.08
	6	Steven P. Abney	2	5.56	11.12
	7	Barry Fishman	2	4.75	9.50
	8	Ronald W. Marx	2	3.99	7.98
	9	Harris Wu	2	3.76	7.52
	10	Phyllis Blumenfeld	2	3.07	6.14
	11	Mark Guzdial	2	2.99	5.98
	12	Michael Curtis	3	1.99	5.97
	13	William Bobrowsky	3	1.99	5.97
	14	John C. Cherniavsky	3	1.83	5.49
	15	June M. Abbas	2	1.99	3.98
	16	Cathleen A. Norris	3	1.00	3.00
	17	J. L. Hoffman	1	1.99	1.99
	18	R. M. Schneider	1	1.99	1.99
	19	J. Poirot	1	1.99	1.99
	20	Paul Oehler	2	0.00	0.00

The table shows the name of the co-author, the total number of collaborations with Elliot Soloway, the RP-Index (RP') of each co-author, and the Co-author Collaboration Value (CCV_a)

best, followed by Pittsburgh, and Maryland, the ranking according to the collaboration activity index shows that Michigan is the best followed by Maryland, and. Syracuse. Berkeley and Pittsburgh are ranked lowest.

The comparison of those two rankings shows that the iSchool of the University of Michigan is ranked highest according to both indices.

Discussion and conclusion

The research questions of “how powerful a researcher’s collaboration network is” and “how much collaboration exists to researchers outside of a community” were answered by defining the *RC-Index* and the *CC-Index*. These indices, which can produce natural numbers or real numbers, were used to assess and compare the iSchools at the University of Pittsburgh, the University of California at Berkeley, the University of Maryland, the University of Michigan, and at Syracuse University.

Our indices consider the quantity of collaborations, the quality of collaborators (by considering the co-author’s productivity), and the number of different co-authors. These indices follow the research of Melin (2000), who showed that 41% of academic survey respondents considered a special competence of a co-author to be the main reason for publishing a joined paper. A high *RC-Index* of a scholar indicates his strong collaboration activities. Equivalently, a high *CC-Index* of a community of researchers shows its strong collaboration activities.

Table 3 The top 10 collaborative researchers of each iSchool

	Name	School ^a	RP'	Co-Au-No	Col-No	RC	RC'
1	Michael Lewis	PITT	2.61	36	119	11	11.99
2	James B. D. Joshi	PITT	5.77	10	28	10	10.39
3	Joseph Kabara	PITT	0.00	18	36	9	9.17
4	K. Sycara	PITT	2.33	19	67	8	8.99
5	Arif. Ghafoor	PITT	5.56	8	25	8	8.99
6	Peter Brusilovsky	PITT	10.99	32	66	8	8.42
7	Jianqiang Wang	PITT	1.99	34	56	8	8.08
8	Prashant Krishnamurthy	PITT	4.40	24	45	7	7.99
9	David Tipper	PITT	4.99	19	43	7	7.99
10	S. Hughes	PITT	2.19	10	30	7	7.99
11	Marti Hearst	UCB	7.99	23	58	13	13.03
12	Marc Davis	UCB	3.68	21	40	10	10.26
13	John Chuang	UCB	4.10	28	55	9	9.53
14	Ka-Ping Yee	UCB	4.99	8	16	8	8.99
15	Kirsten Swearingen	UCB	4.99	9	15	8	8.43
16	Nancy Van House	UCB	4.27	16	30	8	8.38
17	Nathan Good	UCB	5.08	32	43	8	8.19
18	J. Doug Tygar	UCB	5.91	43	53	7	7.99
19	Megan Finn	UCB	2.99	13	22	7	7.99
20	Ariel Schwartz	UCB	4.94	7	19	7	7.99
21	Jimmy Lin	UMD	5.57	49	108	13	13.82
22	Benjamin B. Bederson	UMD	1.37	32	72	11	11.99
23	Boris Katz	UMD	3.99	18	51	11	11.78
24	Dina Demner-Fushman	UMD	4.33	24	48	11	11.20
25	Douglas W. Oard	UMD	5.50	93	180	10	10.80
26	Juan Pablo Hourcade	UMD	1.37	22	47	10	10.56
27	Jianqiang Wang	UMD	0.00	32	52	9	9.33
28	Allison Druin	UMD	7.99	57	126	8	8.99
29	Dagobert Soergel	UMD	2.99	53	84	8	8.99
30	Jennifer J. Preece	UMD	6.11	43	60	8	8.99
31	Elliot Soloway	UMICH	6.99	34	104	18	18.02
32	Joseph Krajcik	UMICH	4.99	29	70	15	15.99
33	Chris Quintana	UMICH	3.99	20	59	15	15.77
34	Mark W. Newman	UMICH	4.80	24	71	14	14.72
35	W. Keith Edwards	UMICH	3.99	17	52	12	12.99
36	Judith S. Olson	UMICH	5.99	28	60	12	12.54
37	Jana Z. Sedivy	UMICH	2.99	12	42	12	12.45
38	Kathleen Luchini	UMICH	3.20	13	37	12	12.04
39	Steven P. Abney	UMICH	5.56	70	97	11	11.99
40	Gary M. Olson	UMICH	5.99	26	48	11	11.99
41	Kevin Crowston	SYR	11.15	28	75	16	16.99
42	Ping Zhang	SYR	9.18	23	45	10	10.10
43	Jeffrey M. Stanton	SYR	4.41	32	52	10	10.02

Table 3 continued

	Name	School ^a	RP'	Co-Au-No	Col-No	RC	RC'
44	James Howison	SYR	7.99	9	27	9	9.99
45	Ian MacInnes	SYR	3.66	21	37	9	9.97
46	Elizabeth D. Liddy	SYR	5.99	39	85	9	9.76
47	Ozgur Yilmazel	SYR	2.08	16	44	9	9.40
48	Barbara H. Kwasnik	SYR	4.99	9	18	9	9.24
49	Anne R. Diekema	SYR	1.62	14	45	8	8.86
50	Robert Heckman	SYR	4.96	10	21	8	8.67

The table shows the name of the researcher, the affiliation of the researcher (School), the researcher productivity index (RP'), the number of different co-authors (Co-Au-No), the total number of collaborations (Col-No), the researcher collaboration index as a natural number (RC) and as a real number (RC')

^a iSchool of University of Pittsburgh (PITT), iSchool of UC Berkeley (UCB), iSchool of University of Maryland (UMD), iSchool of University of Michigan (UMICH), and the iSchool of Syracuse University (SYR)

Table 4 Ranking communities in decreasing order of the CC-index

	School	CC	CC'	CP	CP'
1	Michigan	12	12.83	8	8.41
2	Maryland	9	9.99	6	6.13
3	Syracuse	9	9.88	8	8.01
4	Berkeley	8	8.87	8	8.33
5	Pittsburgh	8	8.62	6	6.97

The table shows the school name (School), the community collaborative index as a natural number (CC), the community collaborative index as a real number (CC'), the community productivity index as a natural number (CP), and the community productivity index as a real number (CP')

While the indices (i.e. RC-Index, CC-Index), which have been introduced in this paper, are more precise, they come with two limitations. First, the calculation of the indices requires substantial input from users than other indices. Second, since our collaboration indices are based on another index (i.e. on researcher/community productivity index such as the h-Index, g-Index, RP-Index, or CP-Index), its calculation requires also the calculation of the productivity index, doubling the effort. Therefore, in order to get a wide acceptance of these indices, the user needs support in calculating them. An application could help which extracts input data (e.g. collaboration relationships) and gives access to the co-author collaboration values and the productivity indices of other researchers.

Despite this practical shortcoming of our indices, our indices (RC-Index and CC-Index) provide a first support for identifying researchers that are suitable for leading research projects, which require integrative work. For this kind of projects, collaboration is an important factor besides the cost factor and the creativity factor. Jiang (2008) showed that these factors are necessary for maximizing research output.

Future work

Additional social context information of authors such as the role of an author (e.g. advisor, student, and colleague) is useful for a more detailed analysis of collaboration. It could be used to extend this research to study scientific collaboration networks following the work of Suresh et al. (2007), which investigated the dynamics of collaborations between students and advisors.

Research indicators, such as the RC-Index and the CC-Index, have been used for various assessments and resource allocation purposes in the past (Dainesi and Pietrobbon 2007). Although the RC-Index and CC-Index are more useful than productivity indices for selecting suitable scientists for heading research projects, the interrelation between our indices and productivity indices is still unclear. In the future, we are going to investigate this interrelation between the RC-Indices (the CC-Indices) and different productivity indices.

Appendix 1. Description and conceptual meaning of the RP-Index and CP-Index

This is an extract from the paper “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), pp 104–118.” However, the indices have been adapted to the notation used within this paper and to address some shortcomings of the real number versions of the indices.

Researcher Productivity Index (RP-Index)

The basis for the Researcher Productivity Index (RP-Index) is the normalized number of paper citations of a researcher j (NC_{ji}). The NC_{ji} is calculated as the number of citations of paper i of researcher j divided by the number of years that the paper is available and multiplied by a factor C_{ji} , which represents the contribution of the researcher j to the paper i (see Eq. 6). The Contribution Factor C_{ji} is in the range between 0 and 1. For example, if all authors of the paper i contributed equally, the factor is $C_{ji} = 1/\text{NumberOfAuthorsOfPaper}_{ji}$. If the researcher j has been a leading author, the factor can be increased. However, the sum of the factors for each author should add up to one.

$$NC_{ji} = \frac{\text{NumberOfPaperCitations}_{ji}}{\text{AgeOfPaper}_{ji}} * C_{ji}, \quad 0 \leq C_{ji} \leq 1. \quad (6)$$

The researcher productivity index (RP-Index) of researcher j is similarly defined to the h-Index and g-Index. Given that the publications of the researcher j are sorted according to the NC_{ji} of each publication i in decreasing order, the RP-Index is defined as the largest natural number x such that the top x publications have at least in average a value of x for their NC_{ji} .

$$RP_j = \max \left(x \mid \frac{1}{x} \sum_{i=1}^x NC_{ji} \geq x \right), \quad x \in \mathbb{N}. \quad (7)$$

If we allow real numbers for the RP-Index, we can obtain a more fine-grained evaluation of the performance of researchers than the h-Index. Since real numbers are used for

evaluating people in many different contexts (e.g. school, college, and graduate school), it is also appropriate to use it for evaluating researchers. Following this approach, we obtain RP' (Eq. 8) after slightly modifying Eq. 7:

$$RP'_j = \begin{cases} \frac{1}{RP_j} \sum_{i=1}^{RP_j} NC_{ji}, & \text{if } \frac{1}{RP_j} \sum_{i=1}^{RP_j} NC_{ji} < RP_j + 1 \\ RP_j + 0.99, & \text{if } \frac{1}{RP_j} \sum_{i=1}^{RP_j} NC_{ji} \geq RP_j + 1 \end{cases}. \quad (8)$$

Community Productivity Index (CP-Index)

Based on the same concept for evaluating the productivity of researchers, a Community Productivity Index (CP-Index) can be introduced. A community can be any group of individuals. The CP-Index of a research community k is defined as the largest natural number x such that the top x researchers of this research community have at least in average a value of x for their RP-Index, given that the researchers are sorted according to their RP-Index (RP') in decreasing order.

$$CP_k = \max \left(x \mid \frac{1}{x} \sum_{j=1}^x RP'_{kj} \geq x \right), \quad x \in \mathbb{N}. \quad (9)$$

RP'_j denotes that researcher j belongs to community k . In the same way as for the researcher productivity index, we can generate a more fine-grained evaluation of the community productivity index by using real numbers. Using Eq. 9, we obtain the following equation for the CP-Index:

$$CP'_k = \begin{cases} \frac{1}{CP_k} \sum_{i=1}^{CP_k} RP'_{ki}, & \text{if } \frac{1}{CP_k} \sum_{i=1}^{CP_k} RP'_{ki} < CP_k + 1 \\ CP_k + 0.99, & \text{if } \frac{1}{CP_k} \sum_{i=1}^{CP_k} RP'_{ki} \geq CP_k + 1 \end{cases}. \quad (10)$$

Acknowledgment This paper was partly funded by Ministry of Knowledge Economy of the Republic of Korea.

References

- Abbasi, A., & Altmann, J. (2009). *AcaSoNet: Social network system for the academic community*. TEMEP Working Paper Series, Seoul National University, South Korea.
- 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.
- Borgman, C., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36, 3–72.
- Braun, T., Glanzel, W., & Schubert, A. (2005). A Hirsch-type index for journals. *The Scientist*, 19(22), 8.
- Dainesi, S. M., & Pietrobon, R. (2007). Scientific indicators of productivity: Time for action. *Revista Brasileira de Psiquiatria [Online]*, 29(2), 100–101.

- Egghe, L. (2006). Theory and practice of the g-Index. *Scientometrics*, 69(1), 131–152.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572.
- 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. (in Chinese).
- 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.
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, 11(13), 3–15.
- Leclerc, M., & Gagn, J. (1994). International Scientific Cooperation: The continentalization of science. *Scientometrics*, 31(3), 261–292.
- Lehmann, S., Jackson, A. D., & Laurup, B. (2006). Measures for measure. *Nature*, 444, 1003.
- Melin, G. (2000). Pragmatism and self-organization research collaboration on the individual level. *Research Policy*, 29(1), 31–40.
- Melin, G., & Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, 36, 363–377.
- Prathap, G. (2006). Hirsch-type indices for ranking institutions' scientific research output. *Current Science*, 91(11), 1439.
- Ruane, F. P., & Tol, R. S. J. (2008). Rational (successive) h-indices: An application to economics in the Republic of Ireland. *Scientometrics*, 75(2), 395–405.
- Schubert, A. (2007). Successive h-indices. *Scientometrics*, 70(1), 201–205.
- Sidiropoulos, A., Katsaros, D., & Manolopoulos, Y. (2007). Generalized h-index for disclosing latent facts in citation networks. *Scientometrics*, 72(2), 253–280.
- Suresh, V., Raghupathy, N., Shekar, B., & Madhavan, C. E. V. (2007). Discovering mentorship information from author collaboration networks. *Lecture notes in computer science. 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, 149–155.
- Van Raan, A. F. J. (2006). Measuring science. In H. F. Moed, W. Glanzel, & U. Schmoch (Eds.), *Handbook of quantitative science and technology research*. Germany: Kluwer Academic Publishers.
- Ziman, J. (1994). *Prometheus bound, science in a dynamic steady state*. Cambridge: Cambridge University Press.