MEASURING THE ACADEMIC IMPACT OF RESEARCHERS BY COMBINED CITATION AND COLLABORATION IMPACT

Jielan Ding [1], Liying Yang [1], Qing Liu [2]

[1] *dingjielan@mail.las.ac.cn, yangly@mail.las.ac.cn*
National Science Library of the Chinese Academy of Science
33 Beisihuan Xilu, Zhongguancun, Beijing, 100190 (P.R.China)

[2] *liuqing@mail.whlib.ac.cn*
The Wuhan Branch of the National Science Library of the Chinese Academy of Science
25 west of xiaohongshan, Wuchang Dist, Wuhan, 430071 (P.R.China).

Abstract
To evaluate the academic impact of researchers in a more comprehensive way, this paper utilizes both the citing and collaborating aspects which are the two main forms occurring in scientific communication. Three citation-based indicators and three collaboration-based indicators are selected and combined into two dimensions by using factor analysis. Then based on the position of the researchers in a two dimensional coordinate system, career roles and career paths of researchers can be revealed. Finally, we provide a theoretical framework for describing the career roles of researchers by the combination of citation impact and collaboration impact.

Introduction
In bibliometrics many studies focus on exploring the academic impact of researchers. It is well accepted that academic impact is produced during scientific communication; therefore measuring the academic impact should be based on the process of scientific communication. Citing and collaboration are the two main forms of scientific communication which can be measured by bibliometric methods. Many published results of investigations are related to citing and collaboration impact.

Most studies concentrated on citation analysis and the indicators used are citation-based, such as total citations, h-index, and citations per publication. This kind of studies stems from the impact in the process of citing, so we refer to this kind of research as "citation impact studies"(e.g. Garfield,1999; Hirsch,2005; Egghe,2006; Jin, Liang&Rousseau,2007; Moed, 2011; Leydesdorff &Bornmann, 2011; Rousseau, 2012).

Another way to detect the impact of researchers, different from using citation indicators is collaboration analysis. With the widely application of social network analysis (SNA) in the field of scientometrics, more and more researchers apply SNA measures to detect academic impact (e.g., Leydesdorff, 2007; Bollen, Van de Sompel, Hagberg &Chute, 2009), especially in collaboration network (e.g.,

Newman 2001a,b; Liu, Bollen, Nelson, & Van de Sompel, 2005; Rodriguez & Pepe, 2008). For researchers' impact study, the micro-level network indicators, such as degree centrality, closeness centrality, and betweenness centrality, are used for measuring researchers' impact in collaboration networks (Yan & Ding, 2009). This kind of research studies the impact produced in collaboration, so we refer to it as "collaboration impact studies".

There is no doubt that citation and collaboration can each describe one aspect of academic impact of researchers. To measure the academic impact of researchers in a more comprehensive way, the two aspects can be combined, in particular as they have a positive relationship (Yan & Ding, 2009; Levitt, Thelwall & Levitt, 2011). A researcher's citation impact is the degree of attention aroused by his academic achievement, so the citation impact represents his academic level to some extent. The collaboration impact reflects one's importance in a certain research community. So, we claim that a combined analysis reflects authors' status in a certain field.

This article aims to obtain both the citation impact and collaboration impact of researchers leading to a more comprehensive way of measuring academic impact. Using a set of indicators based on the two dimensions and factor analysis, we attempt to describe: (1) the career roles of researchers by a combined analysis of the two dimensions; (2) the career paths of researchers by measuring changes in both dimensions.

Methodology

Datasets

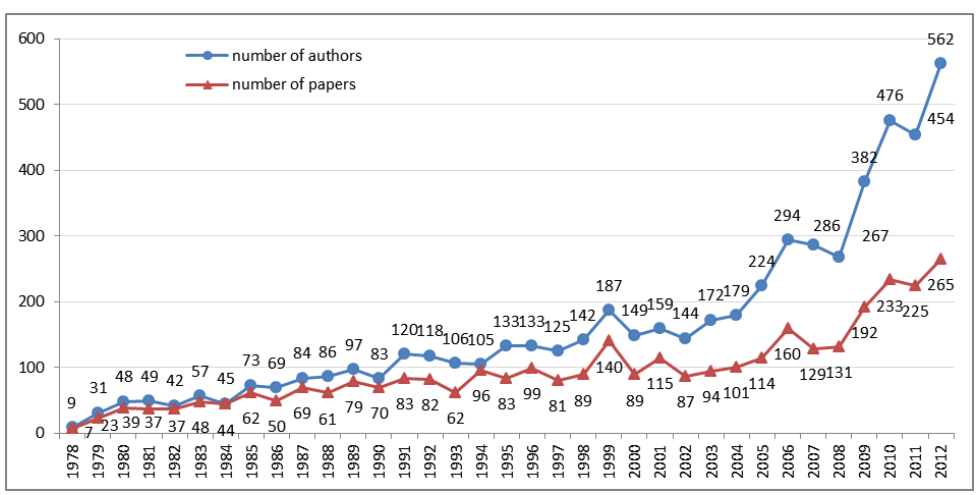


Figure 1. The number of authors and papers of *SCIENTOMETRICS* in 1978-2012. (download date:2013.01.08)

We took the field of scientometrics as an example to apply and test our method. The study focuses on the combined method, and the journal *SCIENTOMETRICS* is the most specialized and typical journal in this field, so the study selected all the papers from *SCIENTOMETRICS* as dataset, though one journal on its own can never describe the entire career of a scientist.

We download the data of *SCIENTOMETRICS* from the Web of Science on 8 January, 2013. The time span is from 1978 to 2012. There are 3376 papers and 3419 different authors after manual data cleaning of authors in the dataset. The yearly data of number of authors and papers of *SCIENTOMETRICS* is shown in Figure 1. Because the increase in the number of papers and authors after the year 2000 is bigger than that before the year 2000, we chose the year 2000 as the cut-off point for detecting the change of impact in the two periods which is the reference year for describing the career paths of the authors. The distribution of authors is shown in Table 1, and the data shows that about 10% authors, who published at least four papers in the period 1987-2012 published over 50% of all papers.

Table1. The distribution of authors.

threshold of authors	*1978-2012 (total dataset)*				*1978-1999 (sub-dataset I)*				*2000-2012 (sub-dataset II)*			
	authors		*papers*		*authors*		*papers*		*authors*		*papers*	
	counts	*%*	*counts*	*%*	*counts*	*%*	*counts*	*%*	*counts*	*%*	*counts*	*%*
published at least 5 papers	223	6.5%	1797	53.2%	81	6.9%	677	47.0%	134	5.4%	911	47.1%
published at least 4 papers	325	9.5%	2018	59.8%	119	10.2%	785	54.5%	210	8.5%	1065	55.0%
published at least 3 papers	495	14.5%	2271	67.3%	175	14.9%	694	48.2%	340	13.8%	1245	64.3%
published at least 2 papers	982	28.7%	2663	78.9%	342	29.2%	1088	75.5%	688	27.9%	1475	76.2%
total	3419	100.0%	3376	100%	1172	100%	1441	100%	2470	100%	1935	100%

Indicators for the two dimensions

A single indicator describes just one aspect of academic impact, but impact is multi-dimensional. Therefore for each dimension we chose a group of indicators from different perspectives to measure the academic impact of researchers.

For the citation impact dimension, total citations, CPP, and h-index were chosen to measure the researchers' citation impact as these are commonly used for academic impact evaluation. We calculated the three indicators based on the citations which the papers published in *SCIENTOMETRICS* received from all papers in WOS.

For collaboration impact dimension, closeness centrality, which means one divided by the total geodesic distance from a node to all others, was chosen to measure the authors' impact over the entire collaboration network, and degree centrality, which means the number of neighbors of a node, was chosen to measure the authors' impact over the local collaboration network. The

collaboration ratio, which means the ratio of collaboration papers among all papers one published, was chosen to measure the depth of collaboration impact. Closeness centrality and degree centrality are network indicators which were calculated in PAJEK based on the *SCIENTOMETRICS* -collaboration network.
We calculated the selected six indicators for each author in the total dataset. These indicators form the base for measuring their career roles. Career paths are based on these six indicators, calculated separately for the sub-dataset I (1978-1999) and the sub-dataset II (2000-2012).

Factor Analysis

Table 2. Rotated Component Matrix[a].

The total dataset of 1978-2012 (KMO: 0.674)			*The sub-dataset of 1980-1999 (KMO: 0.659)*			*The sub-dataset of 2000-2012 (KMO:0.679)*		
	Component			Component			Component	
	1	2		1	2		1	2
total_cites	.927	.135	total_cites	.963	.149	total_cites	.928	.104
CPP	.628	.033	CPP	.544	.171	CPP	.748	-.047
h_index	.936	.166	h_index	.928	.193	h_index	.925	.119
collaboration ratio	-.205	.884	collaboration ratio	-.022	.906	collaboration ratio	-.180	.877
degree centrality	.520	.642	degree centrality	.392	.750	degree centrality	.490	.680
closeness centrality	.359	.622	closeness centrality	.332	.764	closeness centrality	.518	.422

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. a. Rotation converged in 3 iterations.

We chose factor analysis to reduce the six indicators to two integral indicators in order to measure authors' citation and collaboration impact. We did this for two reasons: on the one hand, factor analysis is an objective weighted method of detecting the relationship among a group of indicators; on the other hand, the three citation-based indicators and the three collaboration-based indicators have a positive relationship and their interpretation overlaps to some extent. Therefore these indicators can't be simply added.
For the authors who published at least four papers in the total dataset, we took factor analysis based on their six indicator scores to obtain their citation impact and collaboration impact to describe their career roles. We did the same thing to the authors who published at least four papers in the sub-dataset I (1978-1999) and in the sub-dataset II (2000-2012) in order to detect their career paths.
After factor analysis by SPSS 16.0, the six indicators were reduced to two main components for the total dataset and the two sub-datasets. The results are shown in table 2. A varimax rotation was applied to measure loadings in order to make the components easier to interpret. For the three datasets (sub-datasets), most loadings of the three citation-based indicators are on the first component which we name "citation impact" and most loadings of three collaboration-based

indicators are on the second component which we refer to as "collaboration impact". We note that the two components are orthogonal so there is no linear dependence between the two.

Measuring career roles and career paths for each author

After factor analysis, taking the citation impact component as the abscissa and the collaboration impact component as the ordinate axis, we got each author's location in the Cartesian coordinate system. Each author is located in one of the four quadrants in the two dimensional Cartesian coordinate system. As the four quadrants represent four different career roles we can describe the career role of each author.

To some extent, the ascent and descent of a researcher's citation impact could represent the change of his academic level while the change of a researcher's collaboration impact could represent the change of his activity in collaboration. So the change of authors' impact on the two dimensions can reveal the career paths of researchers.

Results

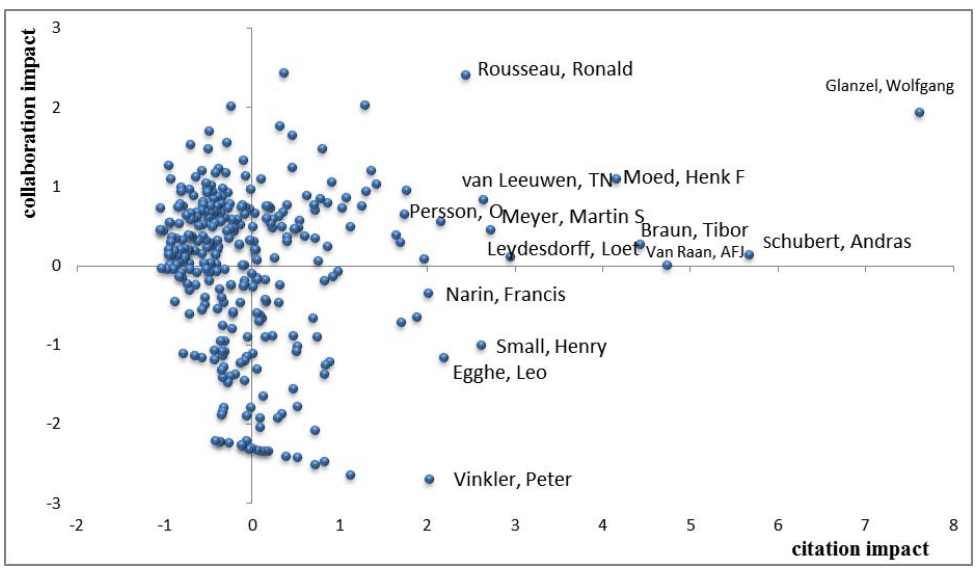
The career roles of researchers in SCIENTOMETRICS

We took the 325 authors in *Scientometrics* who published at least four papers in the period 1978-2012 as examples to describe their career roles by using the two dimensional coordinate system. Then we got the location of each author in the four quadrants which represent four different career roles. The results are shown in figure 2 and the extraordinarily high citation impact authors whose citation impact scores are higher than two are shown on the map. There are 64 authors located in the top right quadrant, 50 authors located in the bottom right quadrant, 139 authors located in the top left quadrant and 72 authors located in the bottom left quadrant.

The authors located in the top right quadrants have both high citation impact and collaboration impact. They not only published highly cited papers which appeal to a lot of people, but also play an important role in scientific collaboration. These authors usually have a very prestige in the field of scientometrics, for example "Glänzel, Wolfgang", "Schubert, Andras", "Van Raan, AFJ", "Braun,Tibor", "Moed, Henk F", "Leydesdorff, Loet", "Meyer, Martin S", "van Leeuwen, Thed N", "Rousseau, Ronald", "Persson, Olle", etc. These high citation and collaboration impact authors are the excellent and core authors.

The authors located in the bottom right quadrant have high cited papers but they rarely collaborate or they are located in the periphery of the collaboration community. Taking "Small, Henry", "Egghe, Leo", "Vinkler, Peter" for example, they are excellent researchers in scientometrics who are Derek John de Solla Price award winners, but they seldom collaborate, especially "Vinkler, Peter" who never collaborated in all his 31

SCIENTOMETRICS papers. These authors of high citation impact but low collaboration are the excellent and lonely (like to work alone) authors.
The authors located in the top left quadrant are very active in collaboration, but relatively ordinary in academic impact. Most internationally oriented Chinese authors who have high prestige in China in the field of scientometrics are placed in this quadrant. Authors of high collaboration impact but low citation impact are the ordinary and core authors.
The authors located in the bottom left quadrant have both low citation and collaboration impact who are the ordinary and lonely authors. They are ordinary in academic achievement and in the periphery of the collaboration network.



(The origin of coordinates is (0,0) which means the average level of the citation and collaboration impact. The authors whose citation impact score are higher than 2 are marked with their names)

Figure 2. The career roles of the 325 authors in *SCIENTOMETRICS* who published no less than 4 papers in year 1978-2012.

The career paths of researchers in SCIENTOMETRICS

We defined seven career paths according to the change of position based on impact score in two periods for each dimension. These are shown in Table 3. There are seven career paths for each dimension; therefore there are 49 combinations representing 49 different possibilities of authors' career status shown in table 4. highly characteristic highly characteristic
We focused on the career path of top authors, and we took the top 20% authors as the top authors for the sub-dataset of year 1978-1999 (first period) and the sub-dataset of year 2000-2012 (second period) for each dimension. There are 96 "top authors" who get into the list of the top authors in citation or collaboration

dimension in at least one period. They form the examples for our career paths analysis. Their career paths are shown in Table 5.

Table 3. The definition of seven career paths.

Description of authors	*career paths*	*the score in first period*	*the score in second period*
top authors	Plateau	>=A	>=B
	New force	—	>=B
	Fall	>=A	—
	Rise	< A	>=B
	Decline	>=A	< B
others	Go up	< A	< B, and higher than the score of former period
	Go down	< A	< B, and lower than the score of former period

(A and B are thresholds for selecting the top 20% authors for each time period which are shown in table 4. "—" means the author didn't have a score because he published less than 4 papers and didn't occur in sample dataset.)

Table 4. Values of A and B for selecting the top 20% authors.

dimensions	*A (threshold for first period)*	*B (threshold for second period)*
citation impact	0.41	0.47
collaboration impact	0.95	0.63

Table 5. The career paths of the 96 top authors (profile).

collaboration impact / *citation impact*	*plateau*	*new force*	*rise*	*go up*	*go down*	*decline*	*fall*
plateau	-	-	2	3	6	4	-
new force	-	6	-	15	-	-	-
rise	-	-	1	2	2	2	-
go up	3	27	2	-	-	1	-
go down	-	-	-	-	-	1	11
decline	-	-	-	1	1	1	-
fall	-	-	-	-	5	-	1

The authors whose career path is "new force" in citation dimension and "going up" in collaboration dimension are likely the "new force" who shifted from other fields to this field. They are absent in citation impact dimension in this field in the first period, but jump to be top researcher by publishing some highly cited papers in the second period. At the same time they have not yet constructed a broad partnership in this field. Porter, Alan L. is such an example (being an expert in data mining and industry-university relations).

The authors whose career path is “new force” both in citation dimension and collaboration dimension are likely the new excellent researchers who grow up in an excellent community of this field. They didn’t occur in the first period dataset, but jumped to be a top researcher in both dimensions in the second period. Liang, Liming is a case in point, collaboration with Rousseau, Ronald.
The authors whose career path is “plateau” in the collaboration dimension and “going up” or “rising” in the citation dimension are located in the center of the collaboration network work in the two periods and become more excellent in academic achievement in the later period, such as “Noyons, ECM”.
The authors whose career path is “plateau” in citation dimension and “rising” or “going up” in collaboration dimension are excellent in academic achievement and become active in collaboration and get to the core position of the collaboration network. Taking “Glänzel, Wolfgang”, “Rousseau, Ronald”, “Leydesdorff, Loet” for example, their citation impact was very high in the two periods and in the second period they get to be located in the central position of the collaboration network.
Most internationally oriented Chinese authors’ career path is “new force” in collaboration dimension and “going up” in citation dimension. Collaboration obviously increases their impact.
The authors who retired and stopped doing research in scientometrics end up in a “going down” or “falling” position. Taking “Narin, Francis” for example, his career path is “falling” in citation dimension and “going down” in collaboration dimension. He established CHI in 1968, an internationally recognized research consultancy company specializing in developing evaluation tools and indicators for science and technology analysis, and obtained the Derek John de Solla Price award in 1988. He retired from CHI in 2004.

Conclusion

Since scientific collaboration becomes more and more popular, it is well accepted that researchers’ citation impact and collaboration impact are equally essential. This investigation took the citing and collaboration dimensions simultaneously into account, leading to a new approach to measure the academic impact of researchers in a more comprehensive way.
By combined analysis of citation impact and collaboration impact, we can discover the detailed information about authors’ career status, such as career roles or career paths. Based on empirical results, we provided a framework to describing the career roles of researchers which is shown in Figure 3. The origin of coordinates means the average level. A researcher’s citation impact represents his academic level. We used “excellent” to describe researchers when their citation impact is higher than the average level and used “ordinary” to describe researchers when their citation impact is lower than the average level. The collaboration impact can usually reflect one’s importance in a research community. We used “core” to describe researchers when their collaboration impact is higher than the average level and used “lonely”(like to work alone) to

describe when their collaboration impact is lower than the average level. Thus we can describe the researchers by using four career roles. If someone has a high score in both citation impact and collaboration impact, he/she is probably an excellent and core researcher. If someone has a high score in citation impact but a low score in collaboration impact, that person is very likely an excellent and lonely researcher in the field. If someone has a high score in collaboration impact but low score in citation impact, he is probably an ordinary and core researcher. If someone has a low score in both citation impact and collaboration impact, he is just an ordinary and lonely researcher (at least at the moment of investigation). For S&T policy makers, identifying researchers in the quadrant we proposed may help them in finding key researchers, possibly with high collaboration (social) skills.

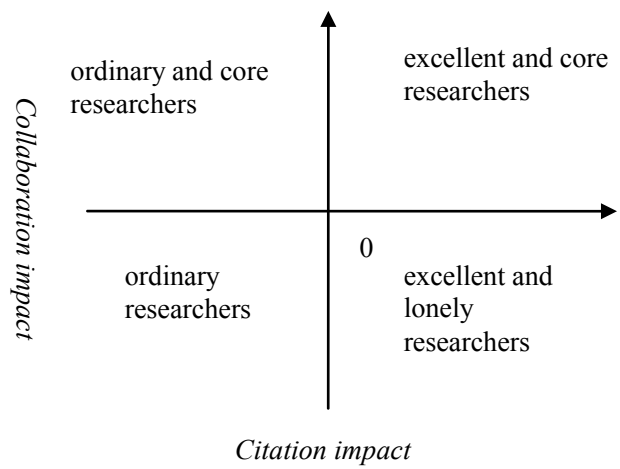


Figure 3: A framework for revealing the career roles of researchers

By detecting the impact changes in two dimensions in different time windows, we could distinguish various career paths for each researcher. The researchers who are top researchers in citation impact for the two periods and become more active in collaboration impact are the "evergreen tree" scientists of this field. The "new force" researchers are those who were not main researchers (they published less than 4 papers) in the first period and turned to be top researchers in the second one. They could perhaps be described as "dark horses" in this field. Finding those different types of researchers is also a way to evaluate researchers from different perspectives.

In empirical study, we detected the career roles and career paths of authors in *SCIENTOMETRICS*. The result may not reflect the real lifetime career roles or paths for researchers as the dataset is limited to one journal, but it certainly provides (partial) information of scientists' profiles active in our field. We use the

combined method in the field of scientometrics and we will do further study of applying the method in other fields to detect the validity of the method.

Acknowledgements

The study was supported by Knowledge Innovation Program of The Chinese Academy of Sciences (Project NO. 10XZNL9). We thank Ronald Rousseau for giving critically valuable advice and making a lot of linguistic corrections. We further thank Prof. Alan Porter (Georgia Tech) and Ting Yue (National Science Library of CAS) for helpful discussions.

References

Garfield,E. (1999). Journal impact factor: A brief review. Canadian Medical Association Journal, 161(8),979–980.

Hirsch, J. E.(2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the USA,102 (46) : 16569- 16572.

Egghe, L.(2006). Theory and Practise of the g-index. Scientometrics, 69(1):131-152.

Jin, B.H.; Liang, L.M. &Rousseau, R.(2007). The R- and AR-indices: Complementing the h-index. CHINESE SCIENCE BULLETIN, 52(6): 855-863.

Moed, H.F.(2011). Measuring contextual citation impact of scientific journals. JOURNAL OF INFORMETRICS,4(3): 265-277.

Leydesdorff,L. &Bornmann,L.(2011). Integrated Impact Indicators Compared With Impact Factors: An Alternative Research Design With Policy Implications. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY,62(11): 2133-2146.

Rousseau, R.(2012). Basic Properties of Both Percentile Rank Scores and the I3 Indicator. JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY,63(2): 416-420.

Bollen, J. , Van de Sompel, H. , Hagberg, A. & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS ONE* ,4(6), e6022.

Levitt, M.J., Thelwall,M., & Levitt, M. (2011). To what extent does the citation advantage of collaboration depend on the citation counting systems? In: (E. Noyons, P. Ngulube & J. Leta, Eds.) *Proceedings of ISSI 2011—13th International Conference of the International Society for Scientometrics and Informetrics, pp.*398–408. Durban: ISSI, Leiden University and University of Zululand.

Leydesdorff, L. (2007)."Betweenness Centrality" as an Indicator of the "Interdisciplinarity" of Scientific Journals. *Journal of the American Society for Information Science and Technology*, 58(9), 1303–1319.

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

Newman, M.E.J. (2001a). Scientific collaboration networks: I. Network construction and fundamental results. *Physical Review E*, 64, 016131.
Newman, M.E.J. (2001b). The structure of scientific collaboration networks. *Proceedings of the National Academy of Science of the United States of America*, 98(2), 404-409.
Rodriguez, M.A. & Pepe, A. (2008). On the relationship between the structural and socioacademic communities of a coauthorship network. *Journal of Informetrics,* 2(3), 195-201.
Yan, E.J. & Ding, Y. (2009). Applying centrality measures to impact analysis: a coauthorship network analysis. *Journal of the American Society for Information Science and Technology*, 10, 2107-2018.