ISSI NEWSLETTER

QUATERLY E-NEWSLETTER OF THE INTERNATIONAL SOCIETY FOR SCIENTOMETRICS AND INFORMETRICS

ISSN 1998-5460

#33 / VOLUME 09 NUMBER 1

MARCH 2013

CONTENTS

IN MEMORIAM:

SUBIR K. SEN

(1947—2013)

BY SILADITYA JANA & RONALD ROUSSEAU

Sri Subir Kumar Sen (9 April, 1947 — 23 January, 2013) was a science writer and researcher. He was attached to the Department of Library and Information Science at the University of Calcutta. From 1991 on (Bangalore, India) he was a regular speaker at ISSI conferences, often also serving as a reviewer of submitted manuscripts. He has published in *Scientometrics*, *Information Processing & Management*, *Current Science* and in the *ISSI Proceedings*. His earlier publications often dealt with Bradford's law and bibliographic scattering. Later he also wrote somewhat more philosophical articles on citing (Sen, 1999) and studied cases of non- or under-citing, concretely work of Shubnikov,

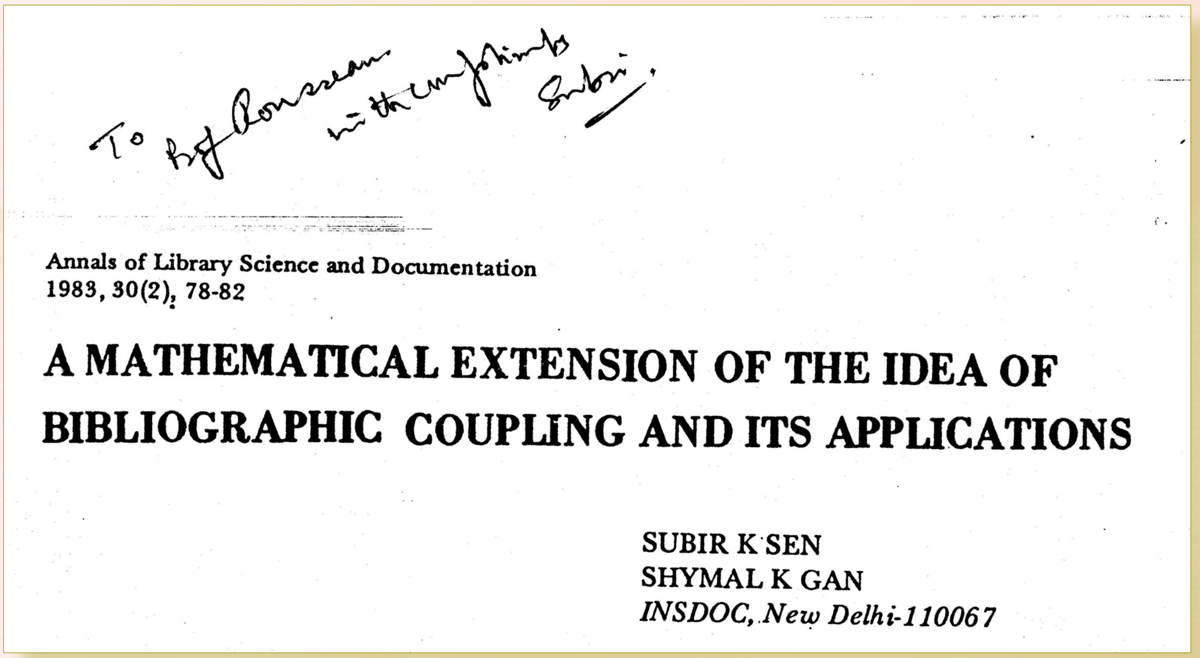
ISSI e-Newsletter (ISSN 1998-5460) is published by ISSI (http://www.issi-society.info/).
Contributors to the newsletter should contact the editorial board by e-mail.

Subir K. Sen (1947—2013)

N. Kumar and K.P. Sinha in superconductivity (Sharma & Sen, 2005, 2006). In his most recent contributions he studied science in Bengal in the 19th century (Roy & Sen, 2010; Sen & Jana, 2011).

For most members of ISSI Subir K. Sen is known as the editor of *JISSI: The International Journal of Scientometrics and Informetrics*, an unfortunately short-lived journal which started publication in 1995. This journal published, among other interesting articles, a selection of papers presented at the Fourth International Conference on Bibliometrics, Informetrics and Scientometrics, Berlin, 1993. In a true spirit of sharing knowledge Subir K. Sen made sure that all articles ever published in *JISSI* are nowadays freely available at the ISSI website.

As an informetrician Subir K. Sen is best known for the introduction of two notions: one being the term bio-bibliometrics, the other being the extension of the notion of bibliographic coupling. The term bio-bibliometrics was introduced in (Sen & Gan, 1990). It refers to a method of establishing functional relationships between a scientist's biography and his/her academic and scientific achievements. In (Sen & Gan, 1983) the authors considered bibliographically coupled and co-cited document sets. Besides considering the documents commonly cited by two articles A and B (the set of documents determining the bibliographic coupling strength of A and B), they also considered the relative bibliographic strength of these two articles. They moreover defined similar notions starting from 3, 4, in general n articles. Studying several articles at the same time, they form a matrix of cited and citing documents, leading to clusters (and cliques) of bibliographically clusters items. They even suggest considering several generations of cited and citing articles, and as such this article was cited in (Hu et al., 2011).



To Prof Rousseau ... Subir

Annals of Library Science and Documentation
1983, 30(2), 78-82

A MATHEMATICAL EXTENSION OF THE IDEA OF BIBLIOGRAPHIC COUPLING AND ITS APPLICATIONS

SUBIR K SEN
SHYMAL K GAN
INSDOC, New Delhi-110067

Fig. 1. First page of (Sen & Gan, 1983) signed by Subir K. Sen

Unknown to most Western colleagues Subir K. Sen was also a prodigious writer of popular and semi-popular articles, sometimes in English, but often in Bengali, including two small books on science for juvenile readers. He was a life member of the Indian Library Association, the Bengal Library Association, the Bangladesh Library Association, the Indian Science Writers' Association and several other professional associations.

Subir K. Sen donated his body to the R. G. Kar Medical College, Kolkata for medical research. The authors hope that this short obituary will keep Sri Subir Kumar Sen's scientific contributions alive.

REFERENCES

XJ. Hu, R. Rousseau, J. Chen (2011). Time series of outgrow indices. *Journal of Informetrics*, 5(3), 413-421.

S.B. Roy & S.K. Sen (2010). Scientific research papers by native Bengali authors during the nineteenth century. *Current Science*, 99(12), 1849-1857.

S.K. Sen (1999). Citation theories: different aspects of the game. In: C.A. Macías-Chapula, (Ed.) *Proceedings of the seventh Conference of the International Society for Scientometrics and Informetrics*. Colima: Universidad de Colima & ISSI, 441-449.

S.K. Sen & S.K. Gan (1983). A mathematical extension of the idea of bibliographic coupling and its applications. *Annals of Library Science and Documentation*, 30(2), 78-82.

S.K. Sen & S.K. Gan (1990). Biobibliometrics: concept and application in the study of productivity of scientists. *International Forum on Information and Documentation*, 15(3), 13-21.

S.K. Sen & S. Jana (2011). Ganot's books on physics in 19th and early 20th century Bengal. *Current Science*, 100(6), 803-804.

H.P. Sharma & S.K. Sen (2005). Non-citing to an Indian paper in superconductivity. *Current Science*, 88(12), 1875-1876.

H.P. Sharma & S.K. Sen (2006). Shubnikov: a case of non-recognition in superconductivity research. *Current Science*, 91(11), 1576-1578.

PROPOSAL FOR A YEARLY ISSI CONFERENCE

The first ISSI conference (be it under a different name) took place in 1987 in Diepenbeek (Belgium). Since then this conference has been organized every two years. Now, more than 25 years later the landscape of our field has thoroughly changed. Nowadays informetrics includes the old subfields of bibliometrics and scientometrics, but also new subfields such as webmetrics and altmetrics.

There is no doubt that our field is growing at an incredible rate. The first issue of this year's Scientometrics (Vol. 94, issue 1) contains 438 pages, and nowadays this journal publishes twelve issues a year. The Journal of Informetrics began in 2007 with an issue of 102 pages (and 356 pages for the whole year), while the first issue of 2013 contained 248 pages. Its sixth volume (2012) contained 728 pages. Someone keeping up with the field of informetrics must read about 1000 pages a month (including contributions in PLoS ONE, Nature and Science).

Should not we, as a society, meet more than just once every two years? Other groups do this already. The Leiden STI conference series has joined forces with ENID and became a yearly event, while also the International Conference on Webometrics, Informetrics and Scientometrics (WIS) & COLLNET Meeting has become a yearly event. The ASIS&T Special Interest Group for Metrics (SIG/MET) hosts regular workshops. And, moreover, there are many local events. One may object that a single person cannot possible attend them all, but why should one?

As president of your society I find it my duty to make sure that your needs are taken care of by our society, that you find the ISSI conferences the most interesting for your research and for your scientific network. We should not expect others to do this for us. Moreover, there are differences in focus between these conferences. True to the definition of informetrics, the ISSI conferences deal with all mathematical aspects of the information sciences. Mathematical modelling is NOT the main focus of an indicator conference, while applications to science policy issues are not the main focus of our conferences. We are very open to social network theorists and colleagues involved in complex networks, and would welcome more of them.

Is there a critical mass of scientists to attend yearly conferences? We never had problems reaching a nice group of attendees (and do not want to become a mega-conference). This was true in the beginning years and since then hundreds of young scientists in Spain, South America, China and Iran have joined our field, while also the number of informetricians (in the broad sense) in Europe, the USA and other countries has increased. Yearly conferences will make it easier to popularize our field over the whole world.

Finally, there is a very practical matter in favour of organising more conferences. Now candidate organizers must make the proposal four, and sometimes even six years in advance. This is not realistic. One cannot make promises about locations, sponsors, not even about the own team, so far ahead. Two years of preparation amply suffices.

As president of ISSI I propose organizing yearly ISSI conferences, starting in 2016, and ask the members of our society for their support. Do you want more value (=conference) for your membership money? Please write to me, the secretary and the members of the board so that we can make an informed decision.

Ronald Rousseau, President ISSI

H-INDEX ON EVERYTHING

OLLE PERSSON

Citations can not only be distributed over authors, but also over several other elements of bibliographic records. For a given set of documents, such as those of the journal Scientometrics, we might ask “Which are the strongest concepts, the most influential countries or authors, and which volumes did produce the strongest impact?”

If we calculate *h*-index for several fields of a record, this is what we can say about Scientometrics. It is mostly about science indicators and journal impact. The greatest impact so far comes from the USA, Netherlands, Hungary and Belgium. The strongest contributing authors are Glänzel W, Schubert A, Braun T, VanRaan AFJ and Moed H, who all have 20 papers cited at least 20 times. Like red wines, some years are better than others, and the rule “the older the better” doesn’t apply.

What		Where		Who		When	
h-index	Keywords	h-index	Countries	h-index	Authors	h-index	Years
33	Science	37	USA	30	Glanzel W	24	2006
31	Indicators	36	Netherlands	25	Schubert A	24	2002
23	Impact	32	Hungary	22	Braun T	22	2003
21	Journals	29	Belgium	21	Vanraan AFJ	22	2001
19	Citation	28	UK	20	Moed HF	22	2004
19	Collaboration	28	Germany	17	Leydesdorff L	21	2007
18	Patterns	21	France	15	Meyer M	20	1996
18	Countries	19	Spain	14	Small H	20	2005
18	Technology	18	Canada	13	Rousseau R	19	1998
16	Citation Analysis	18	People’s R. China	13	Vanleeuwen TN	19	1997
16	Res. Performance	17	Finland	13	Lewison G	19	1994
16	Publication	17	India	13	Vinkler P	19	2000
15	Information	14	Taiwan	12	Nederhof AJ	18	1985
15	Model	14	Australia	12	Persson O	18	2008
15	Scientists	14	Sweden	12	Egghe L	17	1999
15	Performance	13	Israel	12	Thelwall M	16	1993
14	Output	13	Denmark	11	Thijs B	16	1989
14	Innovation	13	Brazil	11	Ho YS	15	1995
13	Ranking	12	Italy	11	Barilan J	15	1991
13	Scient. Literature	11	Switzerland	11	Debackere K	15	1992

Table 1. h-index of what, where, who and when of the journal Scientometrics
Data: Scientometrics as of 2013-02-21 from Web of Science

SCIENTIFIC RESEARCH IN WEST AFRICA: A GLOBAL VIEW (2001-2010)

EUSTACHE MÊGNIGBÊTO
BERSI, Cotonou, Republic of Benin
UA, Information and Library Sciences, Antwerp, Belgium
eustachem@gmail.com

INTRODUCTION

West Africa is one of the five African regions as determined by the African Union. It counts 15 countries; in alphabetical order, they are: Benin, Burkina Faso, Cote d'Ivoire, Cape Verde, the Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Nigeria, Niger, Senegal, Sierra Leone and Togo. Three international languages (French, English and Portuguese) are distinguished in the region as the legacy of the colonisation by France, United Kingdom and Portugal. All the West African countries are together member of the Economic Community of West African States (ECOWAS), a regional economic integration organisation. In the early 2012, the region adopted the ECOWAS Policy on Science and Technology (ECOPOST) that recognized the role of science, technology and innovation in regional integration and life conditions improvement.

Scientometric studies on the whole or part of Africa are very limited, compared to other continents, even though an evident interest has been registered recently (e.g. Adams, King, & Hook, 2010; Boshoff, 2009, 2010; NEPAD Planning and Coordination Agency, 2010; Onyancha & Maluleka, 2011; Tijssen, 2007; Toivanen & Ponomariov, 2011). They found that i) Africa's share to the World output is very negligible, ii) the big science producers in Africa are Egypt to the North, South Africa to the South, Kenya to the East and Nigeria to the West; iii) the big producers also drive collaboration links; therefore, they are the backbone of the scientific collaboration network in Africa and connect African regions and Africa to the World; iv) language, culture, colonial ties and geographical close-up are criteria for collaboration; v) hence, the main African countries' common partners are former colonizers, namely France and United Kingdom; however, USA follows as the third common partner country even though it has no former colony on the Continent; vi) cooperation between African countries is weak; and vii) developed countries are the major non African partners of African countries; iv) countries within the same region collaborate more than they do with countries from another region.

Some regions or countries, like the South African Development Community (SADC) and South Africa, have been steadily explored; others lack examination. In this paper, we intended to give a view of the West African research landscape on the eve of the ECOPOST adoption. We used indicators such as the annual output, languages and types of publications, main partner countries, fields share, citable documents, citations and *h*-index. Web of Science was searched in April 2012 and all the publications co-authored by at least one scientist from any West African country were retrieved from the following databases: *Science Citation Index Expanded* (SCI-EXPANDED), *Conference Proceedings Citation Index- Science* (CPCI-S), *Conference Proceedings Citation Index- Social Science & Humanities* (CPCI-SSH) and *Index Chemicus* (IC). 28,380 records covering the period from 2001 to 2010 were analyzed with the *Analyze Results* function of the Web of Science.

ANALYSES

PRODUCTION

The region's annual production presents an increasing trend and goes from 1,641 in 2001 to 4,617 in 2010. The curve is best fitted by a linear trend of which equation is $y = 361.45t + 850$ where y is the number of papers and t the period of time ($t = 1$ in 2001 and $r^2 = 0.96$) (Figure 1). Therefore, each year, the region produced in average 850 additional papers. The publications are mainly in English (95.5%); French papers come far behind (4.35%); other languages (total less than 0.07%) are German, Japanese, Portuguese, Romanian, and Spanish. Almost all the publications (94%) are citable; however, the annual rate of cited document reported to the total of citable documents decreased progressively from 81% in 2001 to 45% in 2010 for the evident reason that recent publications have fewer opportunities to be cited. The average rate of cited publications over the period is 66.20%. Each cited document received an average of 10.39 citations. The breakdown of the publications per document type gives: *Article* 76.67%, *Meeting Abstract* 10.88%, *Article- Proceedings Paper* 2.57%, *Letter* 2.17%, *Review* 2.33% and *Note* 0.14%. The *h*-index equals 100. The region produced mainly in *Medical and health sciences* (49%), followed by *Natural sciences*

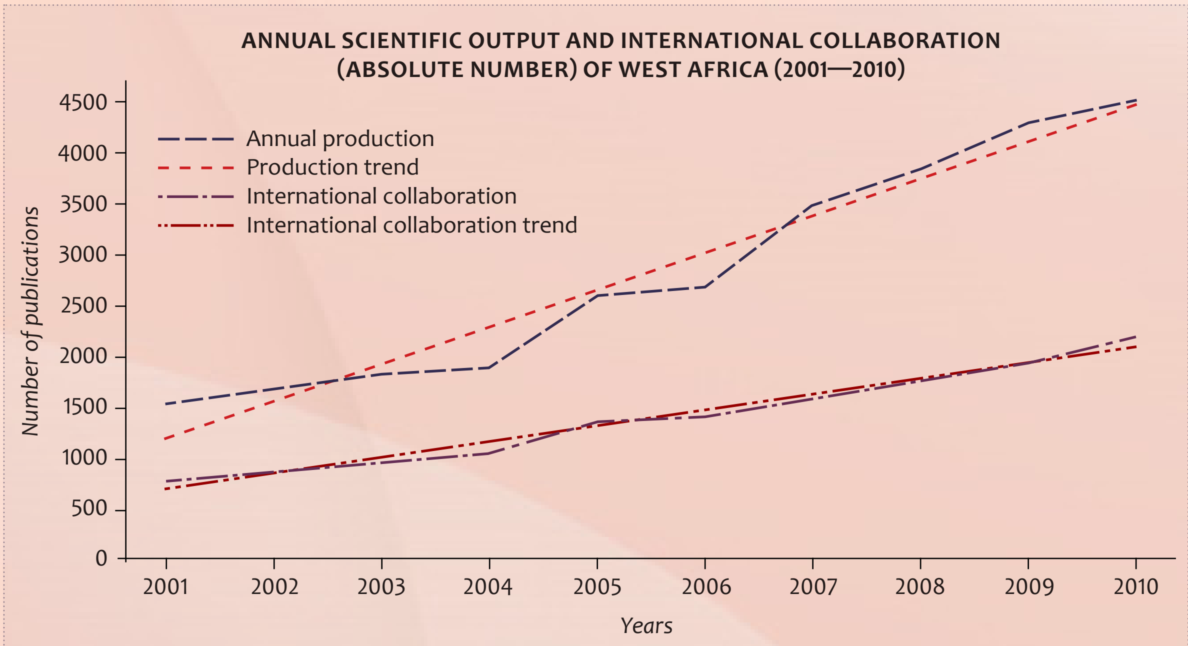


Figure 1 West African scientific output and international collaboration (2001-2010).

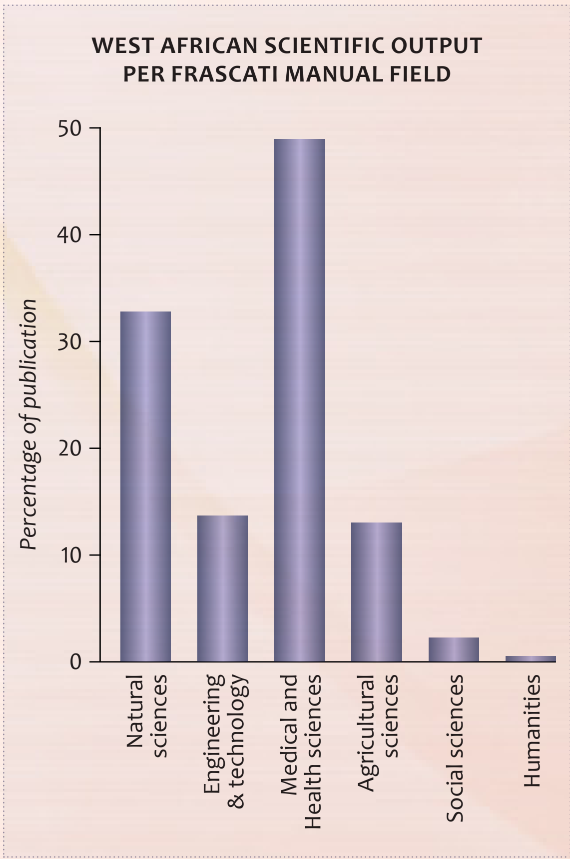


Figure 2 West African scientific output per Frascati Manual Field of Science

	Country	Number of publications	Percentage
1	Nigeria	15,569	54.86
2	Ghana	3,203	11.29
3	Senegal	2,544	8.96
4	Burkina Faso	1,785	6.29
5	Cote d'Ivoire	1,669	5.88
6	Benin	1,335	4.70
7	Mali	1,204	4.24
8	Gambia	986	3.47
9	Niger	586	2.06
10	Togo	433	1.53
11	Guinea	241	0.85
12	Guinea Bissau	225	0.79
13	Sierra Leone	117	0.41
14	Cape Verde	52	0.18
15	Liberia	49	0.17

Table 1 Scientific output of West African countries (2001-2010 - decreasing order)

(32.81%); *Agricultural sciences* and *Engineering and technology* have approximately the same score (around 13%) (Figure 2). Individual countries' production (Table 1) shows that Nigeria is the leader in the region: it is responsible for more than half of the region's annual and total output. This country occupied the same position in the region with regard to criteria like demography (51%), GDP (over 50%), and industrial production (40%) (cf. (Economic Community of West African States, 2010).

INTERNATIONAL COLLABORATION

151 countries contributed to the scientific output of West Africa, among them 38 African, 27 American, 40 Asian, 41 European and 5 Oceanian. They shared 14,094 papers accounting for 49.66% of the total output of the region. The lowest annual international collaboration rate (45.26%) was registered in 2008 and the highest (56.44%) in 2004. The region's top 10 partner countries' shares are computed in Table 2. France is ranked first; it contributed to the West African scientific literature with 3,572 papers accounting for 12.59% of the region's output. Just behind, comes the USA as second partner, with 12.54%, followed by the United Kingdom (10.09%). The fourth partner is Germany, far behind with 4.39%. South Africa is ranked fifth and the first African country with 3%. Out of the top 10 partner countries, 7 are European (France, the United Kingdom, Germany, Belgium, Switzerland, the Netherlands, Italy), 1 African (South Africa) and 2 American (the USA and Canada). The top 10 countries shared with West Africa 12,065 papers accounting for 42.51% of the total output and 85.60% of the papers West Africa shared with the rest of the world. The curve of the absolute number of the West African papers with at least one foreign address follows a linear trend of which equation is $y = 154.73t + 558.4$ ($r^2 = 0.98$), where y is the number of papers shared and t the time, with $t = 1$ in 2001 (Cf. Figure 1).

Europe is by far the first partner of West Africa with 27.12% (9,655 papers) of contribution followed by America 14.55% (4,240 papers), Africa 8.79% (2,530 papers), Asia 5.29% (1,540 papers) and Oceania 1.29% (367 papers).

Rank	Countries	# papers	W. Africa output	World share
1	France	3,572	12.59%	25.35%
2	USA	3,560	12.54%	25.26%
3	UK	2,863	10.09%	20.32%
4	Germany	1,245	4.39%	8.83%
5	South Africa	848	2.99%	6.02%
6	Belgium	830	2.92%	5.89%
7	Switzerland	750	2.64%	5.32%
8	Netherlands	703	2.48%	4.99%
9	Italy	694	2.45%	4.92%
10	Canada	474	1.67%	3.36%
	Total	12,065	42.51%	88.61%

Table 2 West African top 10 partner countries and their shares

At the European continent level, France and the United Kingdom are respectively first and second partner countries with 37% and 29.65% of the number of the West African papers with European addresses. In America, the USA on its own contributed to more than 4 papers out of 5 the continent shares with West Africa. Globally, these 3 countries, with 8,920 papers accounting for 31.43% of the West African total output and 63.28 % of the papers with non West African address, are the main partners of West Africa. At the African continent level, the Southern region is the first partner of West Africa with 1,092 papers (42.11%), followed by Eastern Africa with 890 papers (35.66%), Central Africa with 666 papers (26.88%) and North Africa with (304 papers (12.18%). The main partner countries on the continent are South Africa, Cameroon, Kenya, Tanzania, Uganda and Gabon.

The West African countries' top 5 partners and their shares are computed into Table 3. France is ranked first partner of its former colonies, except Mali; Portugal is ranked first partner for Cape Verde but 13th for Guinea Bissau, the USA in case of Ghana, Liberia, Mali, Nigeria, and Sierra Leone, and the United Kingdom in case of the Gambia only.

COMMENTS AND CONCLUSION

The West African scientific output is trending upwards, which is conformed to the global trend. The region produces mainly in Natural science and Medical and health sciences. The regional annual international collaboration ranges from 45% to 55%. The main partner countries are European or North American at both regional and individual countries' level, France and England, the former colonizers are ranked first and third at regional level

Rank	Benin		Burkina Faso		Cape Verde		Cote d'Ivoire		Gambia	
1	France	31.46	France	35.46	Portugal	42.31	France	44.34	UK	68.05
2	Belgium	13.41	UK	11.20	UK	30.77	USA	13.30	USA	20.79
3	UK	9.36	USA	10.98	Spain	19.23	Switzerland	9.29	Netherlands	7.20
4	USA	9.06	Belgium	8.74	USA	17.31	UK	5.63	France	6.59
5	Netherlands	6.97	Germany	8.57	France	9.62	Belgium	5.57	Belgium	6.19
Rank	Ghana		Guinea		Guinea Bissau		Liberia		Mali	
1	USA	22.35	France	36.51	Denmark	66.67	USA	55.10	USA	40.78
2	UK	20.04	USA	15.35	Sweden	30.22	France	18.37	France	24.17
3	Germany	9.33	UK	14.11	Gambia	21.78	Switzerland	12.24	UK	8.64
4	Netherlands	7.81	Senegal	11.62	UK	16.00	UK	8.16	Switzerland	7.72
5	Switzerland	4.84	Belgium	8.30	USA	8.44	Sierra Leone	6.12	Burkina Faso	7.06
Rank	Niger		Nigeria		Senegal		Sierra Leone		Togo	
1	France	38.40	USA	8.27	France	46.42	USA	29.06	France	30.25
2	USA	18.60	UK	6.19	USA	14.15	UK	23.08	Benin	11.55
3	Nigeria	11.60	South Africa	3.52	UK	6.84	PR China	11.97	Burkina Faso	11.32
4	Burkina Faso	8.53	Germany	2.97	Belgium	5.03	Germany	8.55	USA	10.85
5	Mali	8.53	Italy	1.77	Burkina Faso	4.64	Belgium	7.69	Germany	7.16

Table 3 Top 5 partners of West African countries

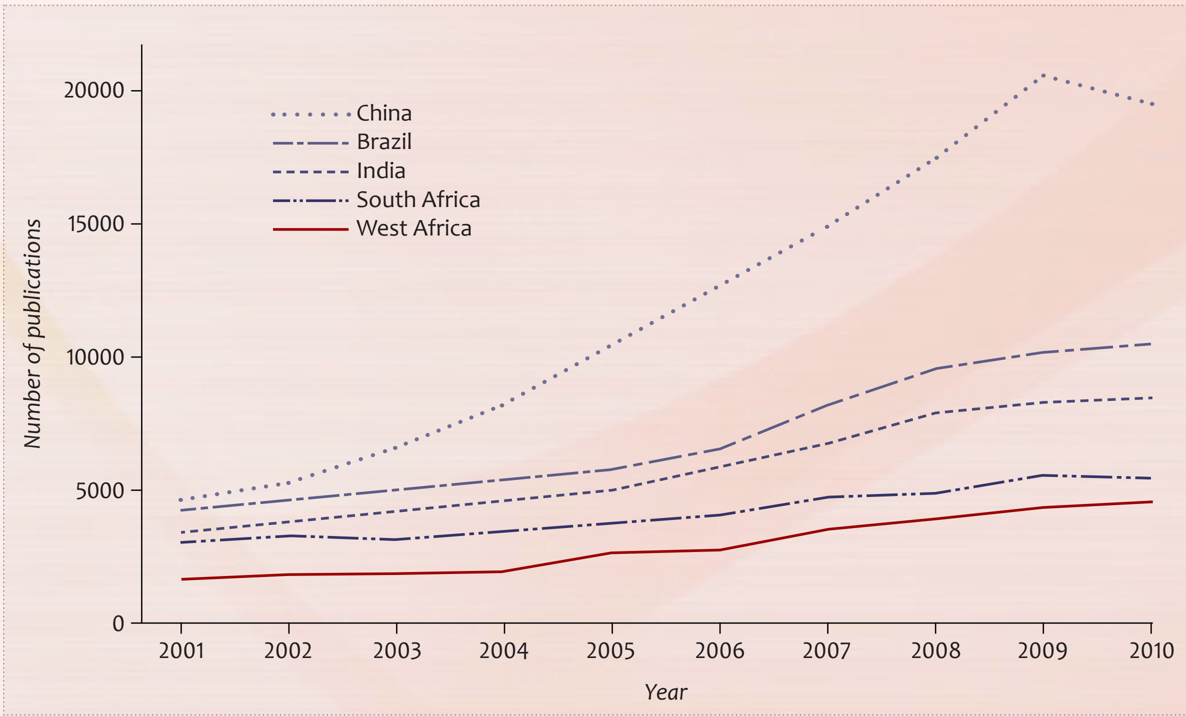


Figure 3 Comparison of the West African scientific output to that of China, South Africa, India and Brazil
Note: For the clearness and the readability of the figure, the Chinese output is divided by 10, the Indian by 6, the Brazilian by 3.75 and the South African by 1.55

	Total output (2001-2010)	Citable (%)	International collaboration (%)	*h*-index	Top 5 partner countries
West Africa (WA)	28,380	94.00	49.66	100	France 12.59 USA 12.54 UK 10.09 Germany 4.39 S. Africa 2.99
Brazil (BR)	261,876	98.28	25.12	285	USA 10.00 France 3.18 Germany 2.77 UK 2.60 Spain 1.82
India (IN)	346,992	97.67	17.56	281	USA 6.06 Germany 2.41 UK 1.76 Japan 1.66 France 1.40
China (CN)	1,199,239	99.23	16.27	350	USA 6.62 Japan 2.15 Germany 1.38 UK 1.35 Canada 1.22
South Africa (SA)	63,087	94.86	43.38	216	USA 14.52 UK 9.12 Germany 5.35 Australia 4.17 France 3.84
Com-parison	WA<SA<BR<IN<CN	WA<SA<IN<BR<CN	CN<IN<BR<SA<WA	WA<SA<IN<BR<CN	

Table 4 Comparing West Africa to China, India, Brazil and South Africa with selected indicators
Note: We computed the West African h-index from the data we collected; China, Brazil, India and South Africa's h-indexes are taken from SCImago (2007) and are related to the period 1996—2011.

confirming Boshoff's (2009) and Toivanen & Ponomariov's (2011) findings. If colonial ties explain this ranking, the position of the USA is a consequence of researchers who have studied in this country and maintained links with former colleagues or supervisors once they have returned back home (Adams et al., 2010). The moderate international collaboration rate at the region's level hides disparities among Member States; indeed, except Nigeria (with 28.42%), individual countries shared 70 to 96% of their scientific production, concurring to Boshoff's (2009) findings. The main partner countries and their shares illustrate the dependence of the regional research system on the Western countries (Toivanen & Ponomariov, 2011). It is the consequence of the weak investment the region and its Member States allot to science, technology and innovation. Boshoff, (2009) reported that Sub-Saharan Africa's countries "are struggling to reach the target of allocating at least 1% of GDP to R & D" as they committed to in the Lagos Action Plan (Organization of African Unity, 1980). In these conditions, science, technology and innovation couldn't get priorities and could not contribute enough to improve population life conditions.

Even though Nigeria, one of the bigger African science producers (Adams et al., 2010) is a West African country, the whole region performs less than Brazil, India, China and South Africa, the leader in science producing on the African continent. Indeed, over the same period of times, according to searches from Web of Science, Brazil produced tenfold the West African volume of papers (63,046), India twelvefold, China about fiftyfold, and South Africa over twofold (Figure 3 and Table 4). Furthermore, the West African *h*-index is lower than that of each of the four countries; therefore, the quality of these countries' research measured by the *h*-index is much higher. Approximately, 43% of South African papers have at least one foreign address; India, China, and Brazil's is lower than 30; hence, West Africa depends more on international collaboration than the four countries. Globally, compared to Brazil, India, China and South Africa, West Africa has the lowest total output, the lowest citable percentage share and the lowest *h*-index; in the opposite, it has the highest international collaboration rate.

REFERENCES

Adams, J., King, C., & Hook, D. (2010). *Global research report: Africa*. Thomson Reuters.

Boshoff, N. (2009). Neo-colonialism and research collaboration in Central Africa. *Scientometrics*, *81*(2), 413–434.

Boshoff, N. (2010). South–South research collaboration of countries in the Southern African Development Community (SADC). *Scientometrics*, *84*, 416–434.

Economic Community of West African States. (2010). *West African common industrial policy – WACIP*. Abuja: Economic Community of West African States.

NEPAD Planning and Coordination Agency. (2010). *African innovation outlook 2010*. Pretoria: NEPAD Planning and Coordination Agency.

Onyancha, O. B., & Maluleka, J. R. (2011). Knowledge production through collaborative research in sub-Saharan Africa: how much do countries contribute to each other's knowledge output and citation impact? *Scientometrics*.

SCImago. (2007). SJR — SCImago Journal & Country Rank. Retrieved from http://www.scimagojr.com

Tijssen, R. J. W. (2007). Africa's contribution to the worldwide research literature: new analytical perspectives, trends and performance indicators. *Scientometrics*, *71*(2), 303–327.

Toivanen, H., & Ponomariov, B. (2011). African regional innovations systems: bibliometric analysis of research collaboration patterns 2005-2009. *Scientometrics*, *88*(2), 471–493.

HOW USEFUL IS THE E-INDEX?

GERHARD J. WOEGINGER
TU Eindhoven, the Netherlands

Suppose that we know the *h*-index and the *g*-index of a researcher. How much additional insight do we gain if we then also learn his *e*-index?

INTRODUCTION

Everybody is familiar with the *h*-index of Jorge Hirsch (2005): *A scientist has index h, if h of his altogether n articles have received at least h citations each, whereas the remaining n–h articles have at most h citations each.* The big plus of the *h*-index are its simplicity and its robustness in quantifying the scientific productivity and the scientific impact of individual researchers. A small minus of the *h*-index is that it ignores citations to highlycited publications: if a publication got cited at least *h* times then it does contribute to the *h*-index, but it does not matter at all whether it got cited exactly *h* times or *h* + 1000 times; these excess citations above the threshold *h* do not have the slightest influence on the *h*-index. As a partial cure for this disease, Leo Egghe (2006a, 2006b) introduced the *g*-index which assigns higher weight to excess citations: *A scientist has index g, if g is the largest integer for which his top g papers together received at least* g^2 *citations.* The *h*-index and the *g*-index are well-known and well-accepted impact indices, and they have received enormous attention from the informetrics and scientometrics community over the last few years.

THE *E*-INDEX

Zhang (2009) introduced the so-called *e*-index for very much the same reasons as Egghe (2006a) did introduce his *g*-index: in order to account for the excess citations that are ignored by the *h*-index. Consider a researcher with *h*-index h, and let $x_1, \ldots, x_h$ denote the number of citations to his *h* top publications. Then *xj–h* of the citations to the j^{th} article are ignored by the *h*-index. The *e*-index simply adds up all these excess citations and then takes the square root from their sum:

$$c^2 = \sum_{j=1}^{h}(x_j - h) = \left(\sum_{j=1}^{h} - x_j\right) - h^2 \qquad (1)$$

Now why is the *e*-index a good indicator for the productivity and the impact of a researcher? Frankly speaking, I do not have the slightest idea how to answer this: the *e*-index is *not* an indicator for impact, and it certainly does *not* measure the productivity of a researcher.

As an example, consider a researcher *X* who has published 100 strong papers, every single one of which has attracted 100 citations. Then *X* has a rightfully high *h*-index of 100 and a rightfully high *g*-index of 100, but a fairly meager *e*-index of 0. In comparison to that consider another researcher *Y* who has published 100 weaker papers, of which 10 each received 20 citations, whereas the remaining 90 papers did not get cited at all. Then researcher *Y* has an *h*-index of 10, a *g*-index of 14, and an *e*-index of 10. Why on earth does the strong researcher *X* have a much lower *e*-index than the mediocre researcher *Y*? The answer is that the *e*-index is lacking most of the good properties that one would expect from a reasonable impact indicator, as for instance monotonicity; see Woeginger (2008a) for a thorough discussion of these issues.

The *e*-index does not express strength or impact of a scientific researcher, and it is somewhat mysterious why it has been baptized to be an "index". The *e*-index cannot stand on its feet alone, and it only makes sense if it is stated in combination with the *h*-index. In other words, the *e*-index is an auxiliary secondary parameter of a citation curve that provides some artificial secondary information. Now of course the question arises: How useful is this secondary information, if we compare it to the information provided by the established, well-accepted, popular and robust *h*-index and *g*-index?

THE *E*-INDEX VERSUS *H*-INDEX AND *G*-INDEX

Let us discuss how the *e*-index relates to the combination of *h*-index and *g*-index. We consider a researcher with *n* publications that have received $x_1, \ldots, x_n$ citations, and we assume without loss of generality that $x_1 \geq x_2 \geq \ldots \geq x_n$. For technical reasons, we furthermore assume that $x_j = 0$ holds for all indices $j \geq n+1$ (this is a standard assumption in the area that avoids tedious range checks for indices). It is well-known that *h*-index and *g*-index always satisfy $g \geq h$, and furthermore the definition of the *h*-index yields $x_h + 1 \leq h$. With this, we derive that the *g*-index satisfies

$$(g+1)^2 > \sum_{j=1}^{g+1} x_j \geq \sum_{j=1}^{h} x_j = h^2 + e^2 \qquad (2)$$

and also

$$g^2 \leq \sum_{j=1}^{g} x_j = \sum_{j=1}^{h} x_j + \sum_{j=h+1}^{g} x_j \leq h^2 + e^2 + (g-h)h \qquad (3)$$

By rewriting the inequalities in (2) and (3) we see that the *e*-index is sandwiched as

$$\sqrt{g^2 - gh} \leq e < \sqrt{(g+1)^2 - h^2} \qquad (4)$$

We stress that the bounds in (4) are well-known in the community and for instance have been stated by Zhang (2010) and Abbas (2012). Now let $r = h/g$ denote the ratio between *h*-index and *g*-index, and note that $0 \leq r \leq 1$. By rewriting and slightly weakening the inequalities in (4) we derive

$$g \cdot \sqrt{1-r} < e \leq g \cdot \sqrt{1 - r^2 + (2g+1)/g^2} \approx \approx g \cdot \sqrt{1-r^2} \qquad (5)$$

In other words, the *e*-index is a priori restricted to an interval whose length is roughly the *g*-index times the multiplicative factor $f(r) = \sqrt{1-r^2} - \sqrt{1-r}$. Since the multiplicative factor *f(r)* is usually small, this interval for the *e*-index is usually very narrow. Elementary calculus shows that $f(r) \leq 0.1683$ for $0 \leq r \leq 1$, and that *f(r)* takes its maximum at $r = 0.6403$; see Figure 1 for a plot of function *f(r)*.

Hence the length of the interval will never exceed the bound of (roughly) *g*/6,

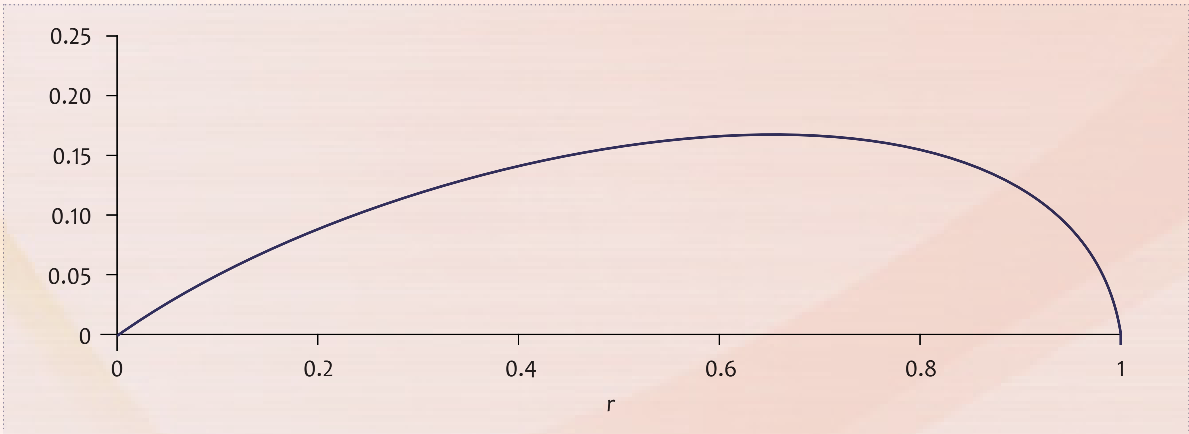


Figure 1: The value $f(r)=\sqrt{1-r^2}-\sqrt{1-r}$ *in terms of the ratio r = h/g.*

and in many cases it will be substantially smaller. Consider for instance the case of a researcher with a certain *h*-index *h* and with a *g*-index $g = 35 \geq h$. Then the *e*-index of this guy is a priori restricted to the interval implicitly defined in (4). The exact borders of this interval of course depend on the value of h, but the length of the interval cannot be larger than 6.

COMPUTATIONAL EXPERIMENTS FOR THE PRICE MEDALISTS

Mozilla Firefox offers a useful add-on for the Firefox web-browser, that has been written by the Italian researcher Agelin Bee (2012): Whenever you query a researcher via Google Scholar, the add-on automatically computes and displays the *h*-index, the *g*-index, and the *e*-index of this researcher. The *h*-index and the *g*-index provide useful information whose meaning is intuitively clear to us. But what does the *e*-index tell us?

We approached this question by looking at seventeen Derek de Solla Price Medallists; see for instance Bar-Ilan (2006) and Egghe (2006b) for previous studies on closely related data sets. We used the Firefox add-on to compute the *h*-index, *g*-index, and *e*-index of these researchers in the areas Engineering, Computer Science, Mathematics and in Business, Administration, Finance, Economics. Our results are summarized in Figure 2. (Note that the add-on always rounds the *e*-index to the nearest integer.)

Now what do we learn from the e-indices in Figure 2? Eugene Garfield has $h = 36$ and $g = 75$, which a priori enforces $55 \leq e \leq 66$. His actual *e*-index is $59 \in [55, 66]$. It is very unclear how to interprete the deeper meaning of this value 59. It lies somewhere in the middle of the interval [55, 66], which means that it could also be somehwat higher or somewhat lower. So what would change, if Garfield's *e*-index would suddenly jump up 61 or even fall down to 55? The numbers seem essentially bare of meaning, and I cannot think of a useful interpretation. Or let us look at Michel Zitt. He has $h = 15$ and $g = 22$, which a priori yields $13 \leq e \leq 17$. Zitt's actual *e*-index is 15 and lies exactly in the middle of the interval [13, 17]. What would change for us, if the add-on would have told us the value $e = 16$ instead?

CONCLUSION

The *g*-index and the *e*-index both have been introduced with the motivation to account for excess citations that are ignored by the *h*-index. Their definitions are closely related, and so are their mathematical behaviors.

Price Medalist	*h*	*g*	*e*	interval
Garfield	36	75	59	55–66
Narin	34	72	59	53–64
Braun	31	54	37	36–45
Leydesdorff	30	54	38	36–46
Van Raan	29	46	30	28–36
Glänzel	28	42	26	25–32
Moed	27	45	32	29–37
Ingwersen	26	60	50	46–55
Small	26	59	50	45–54
Schubert	25	46	33	32–39
Martin	23	52	43	39–47
White	22	49	40	37–44
McCain	21	48	40	36–44
Egghe	21	41	29	29–36
Rousseau	20	36	26	24–31
Vinkler	15	22	15	13–17
Zitt	15	22	15	13–17

Figure 2: The impact indices of 17 Derek de Solla Price Medallists.

The *g*-index is a scientific impact indicator that behaves in a natural way, and satisfies a number of nice properties; see for instance Woeginger (2008b). In particular, high/low *g*-index values indicate high/low scientific impact. In strong contrast to this, the *e*-index does not seem to possess any particularly nice or natural properties. High/low *e*-index values do not correspond to high/low scientific impact, and the only justification of the *e*-index might be that it is easy to compute. All in all, it should be surprising if the *e*-index could stand the test of time.

REFERENCES

A.M. Abbas (2012). Bounds and inequalities relating *h*-index, *g*-index, *e*-index and generalized impact factor: an improvement over existing models. *Public Library of Science ONE 7(4)*, doi:10.1371/journal.pone.0033699.

J. Bar-Ilan (2006). *H*-index for Price medalists revisited. *ISSI Newsletter 2*, 3–5.

A. Bee (2012). Google Scholar *h*-index calculator 2.3.6 (Mozilla add-on for Firefox). https://addons.mozilla.org/en-US/firefox/addon/45283.

L. Egghe (2006a). An improvement of the *h*-index: The *g*-index. *ISSI Newsletter 2*, 8–9.

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

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

G.J. Woeginger (2008a). An axiomatic characterization of the Hirsch-index. *Mathematical Social Sciences 56*, 224–232.

G.J. Woeginger (2008b). An axiomatic analysis of Egghe's *g*-index. *Journal of Informetrics 2*, 364–368.

C.-T. Zhang (2009). The *e*-index, complementing the *h*-index for excess citations. *Public Library of Science ONE 4(5)*, doi:10.1371/journal.pone.0005429.

C.-T. Zhang (2010). Relationship of the *h*-index, *g*-index, and *e*-index. *Journal of the American Society for Information Science and Technology 61*, 625–628.