Jointly published by Akadémiai Kiadó, Budapest *Scientometrics,*
and Springer, Dordrecht *Vol. 64, No. 2 (2005) 121–132*

Scientometrics of computer science research in India and China

SURESH KUMAR, K. C. GARG

National Institute of Science, Technology and Development Studies, New Delhi (India)

An analysis of 2058 papers published by Chinese authors and 2678 papers published by Indian authors in the field of computer science during 1971–2000 indicates that India's output is significantly higher than the Chinese output. However, China is catching up fast. Chinese researchers prefer to publish their research results in domestic journals, while Indian researchers prefer to publish their research results in journals published in the advanced countries of the West. Also the share of papers in journals covered by SCI for India was higher than from China. However, no significant difference has been observed in the impact of the research output of the two countries as seen by different impact indicators. Team research is more common in India as compared to China.

Introduction

India and China are the two most populous countries of the world, which constitutes about 40% of the world's population. Together they represent the world's largest markets. Both countries have very different political and economic systems, but both have recognized the role of information technology (IT) in national development. The governments of both the countries have assigned high priority to IT industry and are emphasizing on expanding the use of IT in schools, public sector agencies and business.

According to a study by Larry Press, et al.,[1] there are an estimated 1.2 million internet connections in China versus only approximately 200,000 in India. China also has a commanding lead in the number of personal computers (PCs). According to one estimate, there are about 25 million PCs in China versus six million in India; and the gap is growing rapidly due to better affordability and a domestic PC manufacturing industry in China. On the other hand, India has a strong base in software industry and is a fast emerging IT superpower. However, its share of earnings is pegged at less than 2% in a $600 billion industry, although it has a potential to get a much bigger slice in the future.[2]

Received November 10, 2004

Address for correspondence:
K. C. GARG
National Institute of Science, Technology and Development Studies
Pusa Gate, K.S. Krishnan Marg, New Delhi 110 012, India
E-mail: gargkc@nistads.res.in

0138–9130/US $ 20.00

It is also important to know the place of these two nations in the world of science. According to a study by Arunachalam[3] India moved down from 8th position in 1980 with 14,983 publications to 15th position in 2000 with 12,127 publications. At the same time, China moved upwards from 15th position in 1980 with 924 publications to 9th position in 2000 with 22,061 publications. However, these included 4307 papers from Hong Kong.[4]

The present study attempts to assess how the two countries perform in research in the field of computer science. The study is based on the analysis of publications indexed in "the collection of computer science bibliographies". Earlier scientometric studies comparing the scientific activity of these two nations in different fields of science and technology have been done by Dhawan[5] for physics, Garg[6] for laser science and technology, and Arunachalam and Gunasekran[7] for diabetes.

Objectives

This study presents a comparative assessment of the status of computer science research in India and China using different scientometric parameters for 1971–2000 (30 years). This period for study has been chosen because there is a steady flow of publications from both countries from 1971 onwards, besides developments in the field of computer science research. Specific objectives of the study are:

(i) To identify the channels used for communication of the research results by the researchers of each of the two countries and to examine the mainstream connectivity based on the pattern of communication;
(ii) To study the relative research effort in computer science by the researchers of each of the two countries and to identify the sub-fields of research where each of the two countries concentrate;
(iii) To assess the impact made by the research output in computer science by each of the two countries using surrogate measures;
(iv) To study the pattern of co-authorship of each of the two countries.

Data and methodology

The data for the present study was compiled from "the collection of computer science bibliographies" available on http://liinwww.ira.uka.de/bibliography/waisbib.html[8] complied by Alf–Christian Achilles. This is a collection of bibliographies of scientific literature in computer science from various sources, covering most aspects of computer science. The bibliographies are updated monthly from their original sources. The collection currently contains more than 1.4 million references (mostly to journal articles, conference papers, and technical reports), clustered in about 1400 bibliographies. The URL provides bibliographies according to the name of the journal, conference, report, or whatever. The compilation of data from the URL involved the following steps:

A. Using "India" or "Indian" as the simple search query option we download a list of all the bibliographies that contain the word "India" or "Indian".

B. All the bibliographies downloaded above in 'A' were individually scanned to get the bibliographic details of each document.

C. Each record downloaded in 'B' above was carefully examined so as to eliminate records that did not truly belong to the search category. Further, each author was searched individually to eliminate non-Indian authors/records.

A similar exercise has been undertaken to download records for China using the term "China" or "Chinese"

Later two databases, one for India and the other for China, were created in FOX-PRO-25 containing details such as record number, document type, author, number of authors, name of the journal, country of publication of the journal, normalized impact factor, and the computer science sub-field for the paper as provided in the bibliography.

Results and discussion

Communication behaviour and mainstream connectivity

The published literature has been classified into journal papers, conferences papers, technical reports, and others. The distribution to literature according of type of literature source is given in Table 1, which indicates that both India and China, like any other country, publish a large portion of their research results as journal articles. However, the share of technical reports published from China is slightly higher than from India.

Table 1. Publication output of India and China by type of literature during 1975–2000

Type of literature sources	India (%)		China (%)	
Journal papers	1855	(69)	1265	(61)
Conference papers	770	(29)	655	(32)
Technical reports	28	(1)	106	(5)
Others	25	(1)	32	(2)
Total	2678	(100)	2058	(100)

The mainstream connectivity of the publication output has been examined by using the characteristics of the journals used for publications, namely the distribution of papers in domestic versus international journals and papers in journals covered by *Science Citation Index* (SCI) versus papers in journals not covered by SCI.

Domestic versus international journals

Apart from native country journals, both Indian and Chinese researchers publish their work often in journals published in the advanced countries of the West such as the USA, the UK, the Netherlands, and Germany. An analysis of the data for the distribution of scientific output in domestic and international journals (Table 2) indicates that the share of publications published in domestic journals by China is about 32 per cent and for papers published in journals from abroad is 68 per cent. For India, the share of papers in domestic journals is 5 per cent and for papers in journals published abroad, the share is about 95 per cent. However, Indian scientists have not published papers in Chinese journals and vice versa. It indicates that Chinese researchers prefer to publish their research results in domestic journals, while Indian researchers prefer more to publish their research results in journals published abroad. One of the possible reasons for less number of papers in international journals by Chinese scientists may be their inability to write articles in English that are of an acceptable standard in terms of English expression. A list of the most commonly used journals used by the researchers of the two countries are given in Appendix 1 and Appendix 2. Of the most commonly used journals by the researchers of the two countries, it is observed that five journals, *Journal of Mathematical Physics* (USA), *Information Processing Letters* (Netherlands), *Applied Mathematics and Computation* (USA), *Theoretical Computer Science* (Netherlands) and *Computer and Graphics* (UK) have often been used by the researchers from both countries. These five journals have 451 papers from China and 563 papers from India, respectively.

Table 2. Distribution of output in domestic and international journals during 1971–2000

Country of publication	India (%)		China (%)	
Domestic	101	(5)	411	(32)
USA	894	(48)	426	(34)
Netherlands	387	(21)	196	(16)
UK	249	(13)	108	(9)
Germany	159	(9)	93	(7)
Singapore	48	(3)	12	(1)
Others	17	(1)	8	(1)
Total	1855	(100)	1265	(100)

From the pattern of communication behaviour, it may be discerned that Indian papers in the field of computer science are better connected to mainstream computer science as compared to the papers published by Chinese scientists. This is because papers that appear in international journals manifest the state of the art in the respective field and, thus, better connected to mainstream science than papers that appear in domestic journals. However, China's preference to publish in domestic journals has resulted in the publication of its own domestic journals in the field of computer science, numbering more than 100; while in case of India, the number of domestic journals published is about only thirty.

SCI versus non-SCI journals

Papers appearing in journals covered by SCI indicate mainstream connectivity and mainstream readership. An analysis of data based on this parameter indicate that about 84 per cent of papers published by Indian scientists appear in SCI journals, while China's share in SCI journals was only 64 per cent. The reason for this may be China's preference to publish in domestic journals, many of which are not covered by SCI.

Relative research effort

This aspect has been examined by using activity index (AI) as the absolute publication output is affected by the size of country as well as the size of sub-field. The activity index (AI) was first suggested by Frame[9] and later elaborated upon by Schubert and Braun.[10] Nagpaul[11] and Garg[12] have used AI in earlier studies. AI characterizes the relative research effort a country devotes to a given sub-field and takes into consideration the effect of the size of country as well as the size of sub-field. AI is the ratio of the country's share of the world's publication output in the given field to the country's share of the world's publication output in all science fields, expressed as percentage. AI = 100 indicates that a country's research effort in the given field corresponds precisely to the world average. AI > 100 reflects higher than average effort, and AI < 100 indicates lower than the average effort by the country.

An analysis of data indicates that Indian researchers published the first computer science paper in 1936, and during 1936 to 1970 they published 69 papers. However, the first paper from Chinese researchers came in 1951, and during 1951 to 1970 they published 13 papers only. This indicates that Indian researchers started working in the field of computer science much earlier than Chinese researchers. Output for India and China for the period 1971–2000 (30 years), in blocks of five years is given in Table 3. During this period China published 2058 (43%) papers, while India published 2678 (57%) papers. This indicates that India's output is significantly higher than China's output.

To compare the relative research effort between the two countries during the period of study, we made the use of AI in a modified way as described below.

Here $AI = \{(Nij / Nio) / (Noj / Noo)\} \times 100$

Nij: Number of publications from India or China in a particular block;

Nio: Total output in that particular block;

Noj: Total output from India or China during 1971–2000; and

Noo: Total output of both countries during 1971–2000.

For example, the AI for India and China for the block year 1971–1975 will be
{(127 /140) / (2678 / 4736) x100 = 160 for India, and
{(13 /140) / (2058 / 4736)}x 100 = 21 for China. Likewise values for other blocks can be calculated for both countries.

The values of the AI during 1971–2000 and the absolute output in blocks of five years are given in Table 3. An analysis of the values of the AI presented in Table 3 indicates that India's activity during the first three blocks (1971–1985) was significantly higher than the Chinese activity. However, from 1986 onwards China caught up with India and during the last three blocks (1986–2000) India's activity index has declined considerably; while that of China has gone up.

Table 3. Publication output and AI from India and China during 1971–2000

Years	India (AI)	China (AI)	Total
1971–1975	127 (160)	13 (21)	140
1976–1980	218 (136)	66 (53)	284
1981–1985	353 (113)	201 (83)	554
1986–1990	590 (98)	473 (102)	1063
1991–1995	774 (87)	793 (116)	1567
1996–2000	616 (97)	512 (104)	1128
Total	2678	2058	4736

AI rounded off to the nearest whole number

Areas of research priority in India and China

The bibliography from which we downloaded the data has classified the bibliographic output in 15 sub-fields. However, sub-fields where the publication output was very small were merged with other sub-fields. Thus, the entire publication output has been classified into eight sub-fields. These are artificial intelligence (AI), compiler technology, programming language and type theory including logic programming, object oriented programming and systems, operating systems and software engineering and formal methods (CT), database research (DR), distributed systems including telecommunication, neural networks and parallel processing (DS), computer graphics and vision (CG), computational mathematics (CM), theory/foundation of computer science (TF), and miscellaneous including typesetting (M).

An analysis of the data on total output in different sub-fields is presented in Table 4. This indicates that emphasis by both India and China on computational mathematics was almost the same. However, Chinese emphasis on compiler technology and programming language as well as theory and foundation of computer science was much higher to Indian emphasis. These three sub-specialties together contributed about 49% and 59% of the total output for India and China, respectively.

Table 4. Areas of research priority in India and China in 1971–1985 and 1986–2000

	India			China			
Subject	1971–1985 (AI)	1986–2000 (AI)	Total (%)	1971–1985 (AI)	1986–2000 (AI)	Total (%)	Grand Total
AI	12 (49)	89(127)	101 (4)	9 (91)	57 (91)	66 (3)	167
DR	13 (79)	51(109)	64 (2)	6 (91)	42(100)	48 (2)	112
CG	59 (87)	188 (98)	247 (9)	29(107)	184(107)	213 (10)	460
CM	300(201)	271 (64)	571 (21)	91(152)	351 (92)	442 (21)	1013
TF	45 (54)	201 (85)	246 (9)	32 (96)	288(136)	320 (15)	566
CT	43 (97)	156(124)	199 (7)	12 (68)	89 (79)	443 (23)	642
DS	57 (45)	373(104)	430 (16)	52(103)	373(116)	101 (5)	531
M	169 (91)	651(123)	820 (31)	49 (66)	394 (83)	425 (21)	1245
Total	698	1980	2678	280	1778	2058	4736

AI rounded off to the nearest whole number.

Data have also been analyzed to examine how the emphasis of the two countries changed during 1986–2000 as compared to 1971–1985 in different sub-fields. To examine this we made use of AI in a modified way as described below.

Here AI = {(Sij / Sio) / (Soj / Soo)}x100

Sij: Number of Indian or Chinese publications in a sub-field in one block;

Sio: Total publications from India and China in the sub-field in both blocks;

Soj: Total number of publications for India or China in one block; and

Soo: Total Indian or Chinese output.

For example, the AI for India and China for the block year 1971–1985 for artificial intelligence will be

{(12 /167) / (698 / 4736)}x 100 = 49 for India, and

{(9 /167) / (280 / 4736)}x 100 = 91. Likewise the values for AI for other sub-fields can be calculated for both countries for both blocks.

The values of the AI for all the sub-fields given in Table 4 indicates that for Indian researchers, the emphasis has gone up for all sub-specialties except computational mathematics. Here the emphasis has gone down significantly. For Chinese researchers, the emphasis in computational mathematics has gone down, but no change is observed in emphasis for artificial intelligence and computer graphics and vision. In the case of other specialties, there is slight increase in emphasis.

Impact of research output

It would have been better to use citations to compare the impact of research for both the countries. However, the data set being so huge, we made use of the surrogate measures to examine the impact. Following surrogate measures have been used to examine the impact of research.

Normalized impact

The impact factor is an indicator of the impact of the journal and depends upon the average rate of citations the articles published in it receive. Since impact factor, as suggested by Garfield, can vary with the discipline and type of journal, use of normalized impact factor has been made. This is so because the papers on computer science have been published in a variety of journals dealing with different subjects such as mathematics, computer science and different branches of engineering. The procedure suggested by Sen[13] and used by Garg and Padhi[14] in their study on laser science and technology has been adopted to calculate the normalized impact factor (NIF).

(NIF)ij = [(GIF)ij / Max (GIF)ij] ×10, where

NIF is the normalized impact factor of the journal i in sub-field j; GIF is the impact factor of journal i in sub-field j, and Max (GIF) is the value of the highest impact factor in the set of journals in the study.

Based on the distribution of papers according to the NIF, it is observed that more than half of the papers from both the countries have appeared in low impact journals. However, the share of papers in medium impact journals is slightly higher for India as compared to China, while China's share in high impact factor journals is higher than India (Table 5). Based on this parameter, one can conclude that the impact of papers published by both India and China do not differ significantly as far as the impact factor of the journals was concerned.

Table 5. Distribution of SCI papers according to Normalized Impact Factor

Range of N.I.F	Number of papers India (%)	China (%)
≤1.00 (Low)	878 (56)	469 (58)
>1.00 ≤2.00 (Medium)	551 (36)	254 (32)
>2 (High)	123 (8)	83 (10)
Total	1552 (100)	806 (100)

Normalized impact per paper (NIMP/paper)

Based on the publication pattern and the normalized impact factor of the journal in which the research results were published, we calculate the normalized impact per paper (NIMP/Paper) for the two countries. This has been calculated by using the formula:

$$\frac{\sum_{i=1}^{n} P_i \sum_{i=1}^{n} F_i}{n}$$

where, P_i denotes the number of papers in the i-th journal, F_i denotes the normalized impact factor of the i-th journal, and n denotes the number of papers. Based on this parameter, it is observed that the NIMP/paper for both the countries is 0.72 (Table 6).

Proportion of papers in high quality journals (PHQ)

The average normalized impact factor of all the journals in which the papers were published is 1.1. We have considered those papers as high quality that have appeared in journals with twice the average normalized impact factor, i.e. >2. Analysis of the data based on this parameter indicates no significance difference in the proportion of high quality papers (Table 6).

Publication Effective Index (PEI)

This measure indicates whether the impact of publications of a country in a research field is commensurate with the publication effort devoted to it. The indicator is the ratio of the proportion of the impact (TNIMP %) to the proportion of publications (TNP %). A value of PEI >1 for a country indicates that the impact of publications is more than the research effort devoted to it for that particular country and vice versa.

TNIMP % = (Total normalized impact for India or China / Total normalized impact for both the countries) ×100. For example TNIMP % for India = (1115×100 / 1692) = 65.89 and for China = (577×100 / 1692) = 34.10

TNP % = (Total Indian or Chinese publications / Total number of publications for both countries). For example TNP % for India = (1552×100 / 2358) = 65.81 and for China = (806×100 / 2358) = 34.18

Hence the value of PEI for India = (65.89 / 65.81) = 1 and for China = (34.10 / 34.18) =1

The values of the publication effective index for both countries indicate that it commensurate with their research effort (Table 6).

Table 6. Impact indicators of the two nations

Country	TNP	TNIMP	NIMP/PAPER	NHQ	PHQ	PEI
India	1552	1115	0.72	123	8	1.00
China	806	577	0.72	83	10	1.00
Total	2358	1692	0.72	206	8.7	

Pattern of authorship

Significant modern research and development can be a collective activity and is often conducted by a group rather than by a single individual. For instance, the share of papers written by authors located in two or more institutions rose from about 33% in 1981 to 50% in 1995, while the total papers rose about by 20%.[15] The distribution of papers according to number of authors is given in Table 7. These papers have been divided into four categories. These are single author papers, two author papers, multi-author and mega-author papers. Multi-author papers included papers with three and four authors, while mega-authored papers included papers with five or more authors. An analysis of data indicates that the proportion of single author papers is greater from China as compared to India. However, the proportion of two author papers, multi-author, and mega-author papers are greater for India as compared to China. This indicates that team research in India is greater than China, unlike the authorship pattern in laser science and technology.[12]

Table 7. Distribution of papers according to number of authors during 1971–2000

# of authors	China (%)*		India (%)*	
Single author	783	(38)	550	(21)
Two author	821	(40)	1326	(50)
Multi-author	441	(21)	759	(28)
Mega-author	13	(1)	43	(2)
Total	2058	(100)	2678	(100)

*Rounded off to the nearest whole number

Summary and conclusion

Major portion of the research results are published in journals. Chinese researchers prefer to publish their research results in domestic journals, while Indian researchers prefer to publish their research results in journals published in the west. China publishes a large number of domestic journals in computer science, while the number of domestic journals for India was much smaller. Indian research output seems better connected to mainstream research compared to China. During the period under study, India's research output is significantly higher than Chinese research output. However, China may catch India in due course of time as in the later period its output has increased significantly. Both countries have emphasized similar sub-fields. Emphasis on computational mathematics has declined during 1986–2000 as compared to 1971–1985 for both countries. However, for other sub-fields activity, as indicated by values of AI, have increased. India's share of papers in SCI covered journal is greater than China's share. However, the impact of research, as seen by the normalized impact/paper, proportion of high quality papers, and the publication effective index does not indicate a significant difference. Indian researchers do more team research, as compared to Chinese researchers as seen from the proportion of two, multi-and mega-author papers.

References

1. L. PRESS, W. A. FOSTER, S. E. GOODMAN, The internet in India and China, at http://www.isoc.org/inet99/proceedings/3a/3a_3.htm (27.10.2004)
2. R. IYER, Asian heavy weight economies: China and India, at http://www.path2usa.com/articles/ita0918.html (27.10.2004)
3. S. ARUNACHALAM, Is science in India on the decline?, *Current Science*, 83 (2002) 107.
4. B. M. GUPTA, K. C. Garg, Is science in India on the decline? A rejoinder, *Current Science*, 83 (2002) 1431.
5. S. M. DHAWAN, Comparative study of physics research in India and China based on INSPEC Physics for 1990 and 1995, *Scientometrics*, 43 (1998) 423.
6. K. C. GARG, Scientometrics of laser research in India and China, *Scientometrics*, 55 (2002) 71.
7. S. ARUNACHALAM, S. GUNASEKRAN, Diabetes research in India and China today: from literature based mapping to health care policy, *Current Science*, 82 (2002) 1086.
8. http://liinwww.ira.uka.de/bibliography/waisbib.html (27.10.2004)
9. J. D. FRAME, Mainstream research in Latin America and Caribbean, *Interciencia*, 2 (1977) 143.
10. A. SCHUBERT, T. BRAUN, Relative indicators and relational charts for comparative assessment of publication output and citation impact, *Scientometrics*, 9 (1986) 281.
11. P. S. NAGPAUL, Contribution of Indian universities to the mainstream scientific literature: A bibliometric assessment, *Scientometrics*, 32 (1995) 11.
12. K. C. GARG, P. PADHI, Scientometric study of laser patent literature, *Scientometrics*, 43 (1998) 443.
13. B. K. SEN, Normalized Impact Factor, *Journal of Documentation*, 48 (1992) 318.
14. K. C. GARG, P. PADHI, Scientometrics of laser research literature as viewed through the *Journal of Current Laser Abstracts*, *Scientometrics*, 45 (1999) 251.
15. NATIONAL SCIENCE BOARD, *Science and Engineering Indicators* 1998, (National Science Foundation, Arlington,VA,1998) pp. 5–43, 5–44, A–310 and Appendix Table 5–52.

Appendix 1
Most common journals where Chinese researchers publish

Journal title with country of publication	# papers	Impact Factor
1. *Journal of Mathematical Physics (USA)*	264	1.008
2. *Information Processing Letters (Netherlands)*	64	0.390
3. *Applied Mathematics and Computation (USA)*	42	0.349
4. *Theoretical Computer Science (Netherlands)*	39	0.417
5. *Mathematics of Computation (USA)*	29	1.210
6. *Journal of Computer Science and Technology (China)*	28	0.000
7. *Numeric Mathematic (Germany)*	26	1.210
8. *Acta Mathematica Sinica (China)*	23	0.324
9. *SIAM Journal of Numerical Analysis (USA)*	23	1.531
10. *Computers and Graphics (UK)*	22	0.700

Appendix 2
Most common journals where Indian researchers publish

Journal title with country of publication	# papers	Impact Factor
1. *Journal of Mathematical Physics (USA)*	264	1.008
2. *Information Processing Letters (Netherlands)*	160	0.390
3. *Lecture Notes in Computer Science (Germany)*	131	0.360
4. *Pattern Recognition Letters (Netherlands)*	53	0.346
5. *Theoretical Computer Science (Netherlands)*	52	0.417
6. *Applied Mathematics and Computation (USA)*	44	0.349
7. *Computers and Graphics (UK)*	43	0.700
8. *Journal of Parallel and Distributed Computing (USA)*	39	0.603
9. *Computer Language (UK)*	27	0.200
10. *Parallel Computing (Netherlands)*	26	0.491
11. *Linear Algebra and its Applications (USA)*	26	0.491
12. *BIT (Netherlands)*	25	0.990
13. *The Computer Journal (UK)*	23	0.398
14. *IEEE Transactions on Software Engineering (USA)*	23	0.940
15. *ACM SIGPLAN Notices (USA)*	22	0.189
16. *SIAM Journal on Applied Mathematics (USA)*	22	1.046