Status of Collaboration in Science and Technology in China as Reflected in Coauthorship

WANG Yan* WU Yishan PAN Yuntao MA Zheng

(Institute of Scientific and Technical Information of China)

ABSTRACT: The coauthorship pattern derived from 1997-2001 data in CSTPCD (Chinese Science and Technology Papers and Citations Database) is analyzed to show the status of science and technology collaboration in China. Four different collaborative types, namely papers coauthored by the authors in the same institution (SI), in different institution located in the same province (SP), in different provinces (DP) of China, and in different countries (or overseas regions) (DC) are discussed. The regional and subject distributions of the coauthored papers as well as the general status of collaboration in science and technology in China are studied. It is concluded that the collaboration in science and technology has been enhanced in China in all the four collaborative types. Different regions have different collaborative pattern corresponding to its development levels of economy, science and technology. The different collaborative pattern in terms of subjects is related to the characteristics of the subject itself.

Key Words: coauthorship, collaboration, science and technology, China

1 Introduction

Collaboration is becoming more and more important in the development of science and technology. At the same time, the more and more convenient ways for data sharing and information exchange provided by information and communication technology enhance both international and domestic science and technology collaboration. The pattern of coauthorship reflects the status of collaboration to some extent. Many researchers have done a lot of studies on collaboration in science and technology by investigating the coauthored papers. Liang et al. (2001) studied the age structure of scientific collaboration in Chinese computer science. Wang and Wu (2001) discussed the status and trend of scientific and technical collaboration between China and Australia. Han and Yang (2000) forecasted the development trend of Chinese

* Contact person. Address: 15 Fuxing Road, Beijing 100038, P. R. China.

Email: wangyan@istic.ac.cn

Yan Wang, Yishan Wu, Yuntao Pan and Zheng Ma, *Status of collaboration in science and technology in China as reflected in coauthorship*, In: Guohua Jiang, Ronald Rousseau, Yishan Wu (Eds), Proceedings of the 9th International Conference on Scientometrics and Informetrics – ISSI 2003, Dalian University of Technology, Dalian, China, 2003, 421-429.

participation in international scientific cooperation by studying the papers published in Science in China. Guo et al. (2000) studied the international science and technology cooperation participated by Chinese using coauthored SCI papers from 1991 to 1998, and found out the main cooperative partners and hot topics. Arunachalam and Doss (2000) used the coauthorship data from the CD-ROM edition of SCI 1998 to analyze international collaboration in science in 11 Asian countries and regions, including India and China. The Research Team on Chinese S&T Papers Statistics and Analysis in the Institute of Scientific and Technical Information of China (ISTIC) has reported both the domestic and international S&T coauthorship annually from 1990 to 2002 (Institute of Scientific and Technical Information of China, 1997, 1998, 1999, 2000, 2001, 2002).

Chinese Science and Technology Papers and Citations Database (CSTPCD), which is used as the official data source for Chinese Science and Technology Statistical Yearbook, is a product of the ISTIC. Its data come from the important Chinese scientific and technical journals, which are called “source journals”. Each paper in CSTPCD had been indexed according to 5 collaborative types: single author (SA), coauthored by the authors in the same institution (SI), coauthored by the authors in different institutions located in the same province (SP), coauthored by the authors in different provinces (DP) of China, coauthored by the authors in different countries (or overseas regions) (DC). The indexing is conducted according to the maximum distance principle. That is to say, each paper was designated to only one collaborative type. When a paper has more than one collaborative types, it is attributed according to the following priority: DC>DP>SP>SI. It is called maximum distance principle. Each paper also has a regional code according to the province where the first author’s affiliation is located. The Classification and Code Disciplines, one of the National Standards of P. R. China, issued in 1992 is used for the subject classification. The subject code of each paper is determined according to the content of the paper as well as the subject of the journal. Because of the complexity of the subject coding, only main subject is involved, and each paper has only one subject code.

The coauthorship pattern derived from 1997-2001 data in CSTPCD is analyzed in this paper to show the status of science and technology collaboration in China.

2 Indicators and Data Processing

The ratio of the number of coauthored papers to the total papers and the number of authors per paper are often used to describe the coauthorship pattern. In order to discuss the collaborative pattern in different regions and subject as well as the general collaborative pattern in China, the following indicators are used in this paper.

N: number of papers;

NA: number of authors per paper;

NC: number of coauthored papers;

PC: proportion of coauthored papers = NC / N

NCrt: number of coauthored papers in region r and collaborative type t;

(r=1, 2 ,…32; t=1, 2, 3, 4, refer to SI, SP, DP, DC respectively)

NCst: number of coauthored papers in subject s and collaborative type t;

(s=1, 2, 3…42; t=1, 2, 3, 4, refer to SI, SP, DP, DC respectively)

PCt: proportion of collaborative type

$$PCrt = \frac{NCrt}{\sum_{t=1}^{4} NCrt}$$

r=1, 2, 3…32. Refer to the proportion of coauthored papers in collaborative type t in region r to all the coauthored papers in the same region.

$$PCst = \frac{NCst}{\sum_{t=1}^{4} NCst}$$

s=1, 2, 3…41. Refer to the proportion of coauthored papers in collaborative type t in subject s to all the coauthored papers in the same subject.

STt: standardized proportion of collaborative type

$$STrt = \frac{PCrt}{\sum_{r=1}^{32} NCrt \Big/ \sum_{r=1}^{32}\sum_{t=1}^{4} NCrt}$$

r=1, 2, 3…32; t=1, 2, 3, 4. Refer to the proportion of collaborative type in region r compared with the weighted average proportion for all the regions.

$$STst = \frac{PCst}{\sum_{s=1}^{41} NCst \Big/ \sum_{s=1}^{41}\sum_{t=1}^{4} NCst}$$

s=1, 2, 3…41; t=1, 2, 3, 4. Refer to the proportion of collaborative type in subject s compared with the weighted average proportion for all the subjects.

3 The General Trend of Chinese S&T Collaboration

The total number of papers in CSTPCD, the number of coauthored papers, the ratio of coauthored papers and the number of authors per paper in 1997-2002 are shown in Table 1. It is found that all of these indicators witness growth year by year.

The average annual growth rate of coauthored papers is 15.22% in this period. From 1997-2001, the proportion of coauthored paper increased from 79.03% to 83.25%, while the number of authors per paper increased from 3.03 to 3.25.

Table 1 The General Indicators of CSTPCD

Year	Number of Total Papers	Number of Coauthored Papers	Ratio of Coauthored Papers (%)	Number of Authors /Paper
1997	120851	95510	79.03	3.03
1998	133341	107989	80.99	3.10
1999	161994	132078	81.53	3.14
2000	179971	150925	83.86	3.21
2001	202179	168324	83.25	3.25

The numbers of coauthored paper in different collaborative types are shown in Figure 1. The number of papers in all the 4 collaborative types increases year by year. All the collaborations have been enhanced in this period.

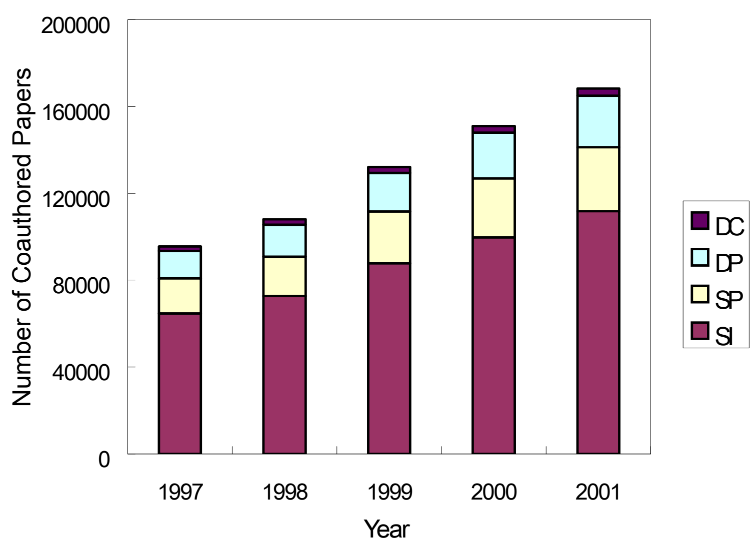

Fig.1 Number of coauthored papers in different collaborative types in 1997-2001

The proportions of the paper amount in each collaborative type in the total number of coauthored papers are calculated, as shown in Table 2.

Table 2 Proportions of different collaborative types (%)

Collaborative Types	1997	1998	1999	2000	2001
SI	67.79	67.28	66.50	66.07	66.43
SP	16.95	16.89	17.97	17.95	17.49
DP	13.16	13.56	13.52	13.99	14.12
DC	2.10	2.27	2.01	1.99	1.96
Total	100.00	100.00	100.00	100.00	100.00

The proportions of SI papers declined while the proportions of DP paper increased year by year. In all the 5 years, the number of papers in different collaborative types follows the same pattern: SI>>SP>DP>DC. It seems that the collaborative priority follows the minimum distance principle. The collaboration is much easier for those in same institution. Zucker et al. (1994) discussed the collaboration structure and information dilemmas in biotechnology, concluded that organizational boundaries act as information envelopes because of trust.

4 Regional Distributions

There were 30 provincial administration regions in the Mainland China in 1997 and 1998. The number has become 31 since 1999 because Chongqing was separated from Sichuan Province. The disparities in the development level of economy, education, culture, and science and technology among these provinces are strong. The status of regional distribution of S&T collaboration is discussed in detail using above indicators.

4.1 General Status of Coauthored Papers in Different Regions

The number of coauthored papers in all the 31 regions is collected for 1997- 2001. The proportions of coauthored paper in all the regions are calculated to show the general picture of S&T collaboration in terms of regional distribution.

The number of coauthored papers increased in all the 31 provinces year by year. Beijing, Shanghai, Jiangsu and Guangdong are the top 4 in terms of number of coauthored papers while they are also the top 4 in terms of total papers. From 1997 to 2001, the average annual growth rate of coauthored papers in the above 4 provinces are 12.76%, 17.83%, 14.74% and 23.41% respectively. The proportions of coauthored papers in all the 31 provinces show the growth trend year by year.

4.2 The Collaborative Preference

The proportion of collaborative type and the standardized proportion of collaborative type are used to show the collaboration preference in different provinces. The data of 2001 is shown in Table 3.

Table 3 Proportion and Standardized Proportion of Collaborative Type (2001)

Region	Proportion of collaborative type (%)				Standardized proportion of collaborative type			
	SI	SP	DP	DC	SI	SP	DP	DC
Beijing	66.96	14.42	15.71	2.91	1.01	0.82	1.11	1.48
Tianjin	70.75	13.58	13.99	1.68	1.06	0.78	0.99	0.86
Hebei	61.85	20.04	16.88	1.23	0.93	1.15	1.20	0.63
Shanxi	64.42	19.82	14.72	1.04	0.97	1.13	1.04	0.53
Inner Mongolia	56.35	25.50	17.37	0.78	0.85	1.46	1.23	0.40
Liaoning	60.98	24.40	12.55	2.08	0.92	1.40	0.89	1.06
Jilin	56.08	25.20	16.32	2.41	0.84	1.44	1.16	1.23
Heilongjiang	61.31	23.57	13.51	1.62	0.92	1.35	0.96	0.83
Shanghai	71.66	13.29	12.60	2.45	1.08	0.76	0.89	1.25
Jiangsu	67.77	16.16	14.04	2.03	1.02	0.92	0.99	1.03
Zhejiang	66.14	19.71	11.55	2.60	1.00	1.13	0.82	1.32
Anhui	63.62	19.40	15.15	1.83	0.96	1.11	1.07	0.93
Fujian	67.83	16.72	13.40	2.05	1.02	0.96	0.95	1.05
Jaingxi	68.16	19.01	12.17	0.65	1.03	1.09	0.86	0.33
Shandong	61.43	23.75	13.44	1.38	0.92	1.36	0.95	0.70
Henan	58.13	24.76	16.45	0.67	0.87	1.42	1.17	0.34
Hubei	68.98	15.84	13.46	1.72	1.04	0.91	0.95	0.88
Hunan	70.01	15.42	13.45	1.11	1.05	0.88	0.95	0.57
Guandong	65.99	17.45	14.25	2.30	0.99	1.00	1.01	1.17
Guangxi	68.44	16.88	13.55	1.13	1.03	0.97	0.96	0.58
Hainan	58.80	12.62	25.91	2.66	0.89	0.72	1.84	1.35
Chongqing	73.98	12.81	11.76	1.45	1.11	0.73	0.83	0.74
Sichuan	68.57	17.45	12.33	1.65	1.03	1.00	0.87	0.84
Guizhou	64.97	17.98	16.47	0.59	0.98	1.03	1.17	0.30
Yunnan	64.86	19.12	14.48	1.53	0.98	1.09	1.03	0.78
Tibet	56.45	12.90	30.65	0.00	0.85	0.74	2.17	0.00
Shaanxi	70.16	14.60	13.62	1.62	1.06	0.84	0.96	0.83
Gansu	60.36	19.91	18.04	1.69	0.91	1.14	1.28	0.86
Qinghai	61.32	15.09	22.17	1.42	0.92	0.86	1.57	0.72
Ningxia	62.73	17.27	19.39	0.61	0.94	0.99	1.37	0.31
Xinjiang	59.80	23.66	16.12	0.42	0.90	1.35	1.14	0.22
Not available*	61.40	16.67	18.42	3.51	0.92	0.95	1.30	1.79
Total	66.43	17.49	14.12	1.96	1.00	1.00	1.00	1.00

*: Means the regional information of the paper not available

The standardized proportion of collaborative type means the collaborative level in one region compared with the average level of all the regions. If this STrt>1, that mean collaborative type t in region r is more active than the average level. It can be seen from Table 3, that Beijing is the most active region in international collaboration, while Chongqing active in the "same institute" type, Inner Mongolia active in the "same region" type, Tibet active in the "different regions" type.

Beijing is the Capital of China, and the seats of many institutes under the Chinese Academy of Sciences and many leading universities in China. It has the advantage in international collaboration. As to Tibet, it is the most remote region in China, and has fewest institutes and universities compared with other regions. So it is the most active region in collaboration with other regions.

It can be seen from the 5 years' data sets that, for most provinces, the pattern of collaborative types is: SI>SP>DP>DC, follows the minimum distance principle. But it is also very interesting to find that, for Ningxia, Qinghai, Tibet and Hainan, which are most remote regions and less developed in terms of GDP, and Beijing and Tianjin, which are two of the four municipalities directly under the Central Government, the pattern of collaborative types is the same, follows the order: SI>DP>SP>DC.

For all the 31 regions, the proportion of SI coauthored papers is obviously higher than other collaborative types.

4.3 The Difference of Collaborative Preference between Regions

The variance analysis is used to investigate the difference of collaborative preference among the 31 regions. The variances of the 5 years' data series of standardized proportion of collaborative type for each region are calculated, as shown in Table 4.

Table 4 Variances of STrt in 1997-2001

STrt of Collaborative Type	1997	1998	1999	2000	2001
SI	0.008	0.008	0.008	0.005	0.005
SP	0.093	0.122	0.057	0.058	0.052
DP	0.074	0.116	0.074	0.065	0.086
DC	0.165	0.165	0.365	0.156	0.135

The difference of variances is consistent for 4 years (except 1998) as: SI<SP<DP<DC. That is to say, the preference of collaboration in the same institution (SI) is common for all the provinces, while the differences in the preference of SP, DP and DC among the 31 provinces are more and more obvious as the collaboration involves longer and longer geographical distance between coauthors.

5. Subject Distribution

The subject distribution of coauthored papers is also investigated.

5.1 General Status of Coauthored Papers in Different Subjects

The number of coauthored papers and the proportion of coauthored papers in each subject are shown in Table 5. They are investigated to find the general picture of collaboration in each subject. The results show that the number of coauthored papers for most subjects is fluctuating year by year from 1997 to 2001, which is obviously different from the result of regional analysis showing that the number of coauthored papers for most provinces increases year by year.

However, the most focused 4 subjects, which claim largest number of papers among all the subjects, namely 1) clinical medicine, 2) computing technology, 3) basic medicine and 4) electronics, communication and automatic control, have the same pattern in terms of the number of coauthored papers, and show steady growth trend in the number of coauthored papers. The average annual growth rate for the number of coauthored papers from 1997 to 2001 in these 4 subjects are 31.36%, 20.65%, 14.09% and 19.69% respectively. And the proportions of coauthored paper in above 4 subjects reached 89.52%, 84.41%, 95.80% and 85.83% respectively.

The top 4 subjects in collaboration are 1) basic medicine, 2) military medicine, 3) chemistry and 4) pharmaceutical science. The proportions of coauthored papers reached 95.80%, 92.87%, 92.20% and 91.13% respectively in 2001. The explanation is that a research team is usually needed to finish a project in these subjects because a lot of experiments are usually needed.

The least collaborative subjects are 1) mathematics, 2) management science, 3) geodesy and cartography and 4) architecture. The proportions of coauthored papers were 60.12%, 62.36%, 67.37% and 69.35% respectively in 2001. Research projects in these fields can be finished by individuals more easily.

Table 5 Number and Proportion of Coauthored Papers in Different Subjects

Subjects	Number of Coauthored Paper					Proportion of Coauthored Paper (%)				
	1997	1998	1999	2000	2001	1997	1998	1999	2000	2001
Mathematics	2218	2445	2474	3128	3146	49.13	50.22	53.12	59.43	60.12
Mechanics	1837	1500	1631	1658	1580	79.59	83.01	83.38	84.76	85.41
Information science & system science	450	775	642	254	1318	69.55	76.51	78.20	76.05	83.31
Physics	4181	3925	4854	4825	4018	84.06	85.81	86.25	87.25	86.88
Chemistry	7054	6223	6351	6921	7475	91.79	92.55	93.58	93.20	92.20
Astronomy	275	315	379	363	426	72.94	70.63	74.90	81.21	78.31
Earth science	4949	5045	5372	5810	5716	77.32	78.53	82.57	83.02	84.29
Biology	5373	5719	6637	8007	8784	88.07	88.80	90.61	91.24	90.98
Prevention Medicine and health	1868	1941	2563	2410	3758	89.94	91.38	85.01	87.00	84.83
Basic medicine	6013	7717	9420	8417	10189	96.07	96.52	95.74	95.21	95.80
Pharmaceutical science	1274	1854	2894	3665	4222	89.53	89.74	90.55	92.29	91.13
Clinical medicine	10381	12497	18466	26965	30906	92.32	92.24	88.94	88.87	89.52
Traditional Chinese medicine	2558	3008	2695	3333	4482	77.92	81.36	78.23	80.31	79.86
Military medicine	172	214	287	231	469	92.97	93.45	90.25	97.47	92.87
Agriculture	4998	5310	6085	6010	6668	83.26	85.52	85.18	87.25	86.90
Forestry	905	1106	1480	1469	1726	82.72	80.38	83.52	83.89	82.23
Animal husbandry and veterinary	1374	1129	1687	1528	1274	84.66	86.51	82.33	87.26	84.54
Fishery	452	480	673	785	687	74.46	75.71	76.83	87.03	83.17
Geodesy and cartography	40	183	167	197	318	64.52	70.93	74.55	88.74	67.37
Material science	3627	2531	3849	3075	2676	84.80	88.25	85.82	88.92	73.76
Basic disciplines in engineering & technology	362	1097	661	880	1231	71.12	71.70	78.22	79.35	74.43
Mining engineering	964	1451	2290	1958	1637	63.01	65.83	67.06	71.54	74.82
Energy science	1848	1947	3460	3391	3055	70.59	73.58	75.83	84.58	80.56
Metallurgy	1168	2477	2613	4779	5655	70.40	80.19	75.70	78.14	75.54
Machine and instrument	4319	5137	5963	4134	4564	70.97	73.99	73.51	74.41	74.39
Power engineering and electrical engineering	3043	3636	4149	4634	5366	71.97	74.52	79.00	79.98	78.60
Nuclear science	582	693	598	719	1020	84.23	85.14	87.55	95.11	90.67
Electronics, communication & automatic control	4800	6976	6404	6921	9852	80.17	83.03	83.80	87.27	85.83
Computing technology	5422	5387	8194	11542	11490	77.28	78.44	81.62	86.84	84.41
Chemical engineering	3187	4824	4065	6199	6092	73.62	77.96	77.77	81.10	79.31
Light industry and textiles	526	721	1422	2480	2080	64.30	68.54	67.04	71.33	71.95
Food science	717	649	841	105	586	75.71	76.26	79.26	86.78	77.72
Architecture	2358	2605	3373	4414	4616	53.54	60.12	62.09	66.03	69.35
Water conservancy	792	942	1251	1687	1470	69.05	74.11	68.55	84.43	71.99
Transportation	1384	1471	2041	2215	2689	62.57	69.88	70.23	69.26	73.27
Aviation and space	1087	654	1193	1378	708	79.69	81.44	85.34	85.22	90.31
Safety science & technology	100	113	64	82	110	64.52	68.07	66.67	64.57	88.00
Environmental science	2224	2348	3600	2889	3925	77.46	78.84	80.84	82.10	84.90
Management science	538	870	957	1080	1115	55.64	61.10	61.66	69.32	62.36
Others	90	74	333	387	1225	53.89	39.36	54.77	53.97	63.01
Total	95510	107989	132078	150925	168324	79.03	80.99	81.53	83.86	83.25

5.2 Subjects Clusters according to the Coauthored Papers

Cluster analysis is used for the classification of the 40 subjects, by examining 10 data sets, which are the number of coauthored papers and proportion of coauthored papers from 1997 to 2001. The 40 subjects can be classified into 4 clusters according to the result. The first cluster only contains clinical medicine, which has highest number of coauthored papers and highest proportion of coauthored papers at the same time. The second cluster contains computing technology, basic medicine, electronics, biology, chemistry, physics, etc, which have large number of coauthored papers and relatively large proportion of coauthored papers. The third cluster contains mining technology, light industry and textiles, safety science and technology, and management science, with low number of coauthored papers and low proportion of coauthored papers. The other subjects belong to the fourth cluster.

5.3 Collaborative Preference in terms of Subjects

For all the 40 subjects, the proportions of SI coauthored papers are higher than other collaborative types. The variances of the standardized proportion of each collaborative type for different subject are calculated, as shown in Table 5.

Table 5. Variances of STst in 1997-2001

STst of Collaborative Type	1997	1998	1999	2000	2001
SI	0.010	0.009	0.010	0.010	0.010
SP	0.134	0.127	0.114	0.118	0.078
DP	0.082	0.068	0.083	0.080	0.096
DC	0.478	0.304	0.484	0.472	0.278

The order of variances basically follows SI<DP<SP<DC.

6 Conclusion

It can be concluded that the collaboration in science and technology are being paid more and more attention in China. The science and technology collaborations in all the 31 provinces are enhanced. The regions with higher number of papers (which can be called active regions in science and technology development) are also the regions paying more attention to science and technology collaboration. All the provinces emphasize the collaboration in the same institution while the preference for collaboration between different institutions, different provinces, and even different countries (or overseas regions) are different from one province to another. Collaboration preference is generally following such an order: SI>SP>DP>DC, which indicates that the geographical distance is still one of the major obstacles for S&T collaboration. Different region has different collaborative pattern corresponding to its development levels in economy, science and technology. The collaborations in focused subjects are enhanced while the collaborations in some subjects are weakened. The different collaborative pattern in terms of subjects is related to the characteristics of the subject itself.

Acknowledgement

This paper is sponsored by the National Natural Science Foundation of China (NSFC). Grant Number is 70073007.

References

Arunachalam S., & Doss M.J. (2000). Mapping international collaboration in science in Asia through coauthorship analysis [on-line].
Available: http://www.collnet.de/workshop/arunachalam.html

Guo J., Cui L., Zhang H., Han D. & Chen J. (2000), Bibliometrics study of international cooperation in science and technology. Journal of the China Society for Scientific and Technical Information, 19(6), 659-662 (in Chinese)

Han X., & Yang N., (2000). An examination of the development trend of Chinese participation in international scientific cooperation from the international co-authorship phenomenon. Information Science, 18(10), 920-926 (in Chinese)

Institute of Scientific and Technical Information of China, Chinese Scientific and Technical Papers Statistics and Analysis (Annual Report) 1997, 1998, 1999, 2000, 2001, 2002 (in Chinese)

Liang, L.M., Kretschmer, H., Guo, Y.Z., & Beaver, D. D. (2001), Age structure of two-dimensional and three-dimensional research collaboration in Chinese computer science. In Havemann, F., Wagner, D. R., & Kretschmer, H. (Eds.), Proceedings of the Second Berlin Workshop on Scientometrics and Informetrics: Collaboration in Science and in Technology (pp. 143-154). Berlin: Gesellschaft fur Wissenschaftsforschung.

Wang, Y., & Wu, Y., (2001). Status and trend of scientific and technical collaboration between People's Republic of China and Commonwealth of Australia: an analysis of scientific and technical papers co-authored by Chinese and Australian. In Havemann, F., Wagner, D. R., & Kretschmer, H. (Eds.), Proceedings of the Second Berlin Workshop on Scientometrics and Informetrics: Collaboration in Science and in Technology (pp. 211-218). Berlin: Gesellschaft fur Wissenschaftsforschung.

Zucher, L.G., Brewer, M.B., Darby M.R., & Peng, Y. (1994). Collaboration structure and information dilemmas in biotechnology: organizational boundaries as trust production. Institute for Social Science Research. Volume VI, 1994-95, Biotechnology Studies. Working Paper 2 [On-line]. Available: http://repositories.cdlib.org/issr/volume6/2

A revised version of this contribution has been published as:

Yan Wang, Yishan Wu, Yuntao Pan, Zheng Ma, Ronald Rousseau (2005). Scientific Collaboration in China as Reflected in Co-authorship. *Scientometrics*, 62, 183-198.