Trends in the inter-regional and international research collaboration of the PRC's regions: 2000-2015

Marc Luwel*, Erik van Wijk*, Lambertus J. Van der Wurff* and Lili Wang**

*{luwel,wijk,wurff}@cwts.leidenuniv.nl

Centre for Science and Technology Studies (CWTS), Leiden University,
P.O. Box 905, 2300 AX Leiden (Netherlands)

**wang@merit.unu.edu

UNU-MERIT, Keizer Karelplein 19, 6211 TC Maastricht (Netherlands)

Abstract

The regional structure of the PRC's scientific output is analyzed using publications processed for the Web of Science. Over the period 2000-2015 and measured by the Salton Index of the co-publications the scientific collaboration among the PRC's regions increased only slightly, in stark contrast with the USA' states and during the most recent years the EU member countries. Only for research with other nations, representing about 30% of the total publication output, inter-regional collaboration is on the rise. For the leading PRC's regions the USA is the dominant partner co-authoring about 50% of their publications. Germany and especially Japan seems to lose attractiveness to the advantage of the UK, Australia and neighboring Asian countries.

Introduction

The People's Republic of China (PRC) became the third economic power measured in nominal GDP only preceded by the USA and the European Union; the latter not being a nation it is now the second wealthiest country. In parallel its science and innovation capacity developed rapidly and measured by the number of scientific publications it is also the world's second most productive nation, again after the USA (National Science Board, 2014)

It is well known that over the last half century the business of science has become more internationally oriented and cross-border collaborations are on the rise especially for natural and life science and the basic disciplines of engineering (Waltman, 2011). Within larger countries intra-regional collaborations are also increasing (Bellini, 2013). To study these phenomena often publications are used as a proxy for a country's scientific capacity.

The PRC's international co- publications have been intensively studied (Glänzel, 2007; Haustein, 2011), often focusing on specific domains (Tang, 2011) or countries (Wagner, 2015). Far less attention has been paid to the PRC's domestic scientific collaboration and the international collaboration patterns of its geographical regions (Liang, 2002; Scherngell, 2011; Sun, 2015; Andersson, 2014).

Research questions, methodology and data

In the PRC there are 31 administrative mainland regions, including 23 provinces, 5 autonomous regions and 4 municipalities, further called 'regions'.

In this paper the regions are the unit of analysis and both the evolution of inter-regional and international collaborations are studied. Trends in inter-regional collaborations are benchmarked with those of the USA and the European Union. The regions' profiles based on the countries they are collaborating with are compared and changes over time analyzed.

Using the Web of Science (WoS), an international bibliographic database produced by Thompson Reuters that covers very well the above mentioned disciplines, information on all publications with the PRC in the byline was extracted. The search was restricted to the period 2000-2015 and to the publication types Articles, Letters and Reviews.

Generally the PRC addresses processed for the WoS contain information on the postal code of the institute's city; the first two or three characters of these ZIP codes indicate the region. Using this information about 85% of the addresses can be assigned automatically to a region. In the remaining addresses the postal codes are missing or erroneous. A combination of manual data cleaning and algorithms based on recurring patterns in the errors allowed to increase the yield to about 95%. In this analysis the remaining 5% of the PRC's publications that could not be assigned to a region, are not taken into account.

For each year and each region the publications are divided into two subsets: those with an address from another country in the byline (further called international co-publications) and those with only one or more PRC addresses (further called domestic publications). In the first subset publications may be signed by authors from more than one region; domestic publications can also have addresses from two or more regions, further called domestic co-publications.

For the domestic publications and for each year a symmetrical co-publication matrix is calculated with on the diagonal the number of domestic publications of each region and on the off-diagonal entries the number of co-publications between two regions. A full or integer counting scheme is used at the level of the regions assigning a co-publication fully to each contribution unit. It should be emphasized that a publication with one or more addresses from only one region is classified as a domestic publication and assigned only once to that region.

For the international co-publications a similar matrix is calculated. For each region and for each year the list of countries in the byline of the publications and their number of co-publications is extracted from the WoS; again a full counting scheme is used

To analyze the inter-regional collaboration the absolute number of publications are an indicator. It is however well known that regions' propensity to collaborate depends on their total number of publications. Similarity measures take this effect into account (van Eck, 2009).

In this paper to quantify the collaboration strengths between the regions the Salton Index (SI) is calculated:

$$Sij(t) = \frac{Pij(t)}{\sqrt{Pi(t) * Pj(t)}}$$

where i and j represent the regions, Pij(t) the co-publications between i and j and Pi(t) the number of publications of region i; t being the publication year (Luukkonen et al., 1993). The SI is a symmetrical matrix with 1 on the diagonal.

To make the evolution of the collaboration strength between the regions visible for each year the average, the median, the maximum and the minimum value of the SI is used as an indicator.

At the same time as the PRC's science and innovation system rapidly developed, other countries' inter-regional collaboration expanded. For the European Union stimulating collaboration between member states is even a long-standing policy objective. To benchmark PRC's inter-regional collaboration its SI is compared with this measure calculated for the first 15 countries joining the European Union (EU15) and for the states within the United States (Luwel, 2015).

One of the factors influencing co-publication activities among regions is not only the available scientific capacity but also their geographical locations. To test for effects of the geographical proximity on the collaborative strength between regions a symmetrical distance matrix was calculated using the geo-coordinates of the capital city of the regions. To test for correlation the Mantel test between the co-publication matrix and the distance matrix on the one hand and the SI and the latter on the other hand is done using the pearson correlation and treating the diagonal elements in the matrices as missing values (Mantel, 1967). This operation was carried out for each year and for both subsets.

Next each region's international co-publication profile is constructed based on a country's number of co-publications with that region and on the ratio between this number and the region's total number of international co-publications. The regional profiles are compared as well as their evolution during the period 2000-2015.

Results

Table 1 gives for the 31 regions the number of international co-publications and domestic publications. The distribution of these publications is highly skewed with the top-10 regions producing more than 75% of the total output. Over the 15 year period the total number of publications (i.e. the sum of the number of domestic and internationally co-authored publications) increased by a factor 10; this increase is roughly the same for the top 10 and the next 10 regions and even higher for the third tier.

In 2015 the ratio of the international co-publications and the domestic publications is about 30%, an increase by 7% compared to 2000; for the top-10 regions this increase is roughly the same.

However table 1 shows that the growth rate decreases slightly over time but the increase in internationally co-authored publications is outperforming the domestic publications in the last two 5 year periods by 44% and 25%.

Next the strength of the inter-regional collaborations and its evolution are analyzed by calculating for each year the SI. Figure 1a shows the evolution of the mean value of the SI for the international and the domestic co-publications for the 31 regions and for the top-10 regions. For the domestic co-publications the average value of the SI remains roughly constant over the period 2000-2015. It is not surprising that for all the 31 regions together the values of the indicator are lower than for the top-10 as the third tier regions have a low number of (co-)publications.

At the beginning of the period the average values of the SI for the international co-publications is below the values for the domestic publications, subsequently during the next few years they have a rather erratic behavior. In the beginning of the last decade most regions even among the top-10, had very few international co-publications; the subset of these publications with addresses from than one region was even smaller. The upswing in the values for the year 2004 could be explained by the incorporation of additional Chinese journals in the WoS and will be investigated in more detail. From 2008-2009 onwards the average values of the SI for the international co-publications increase systematically to a level well above the values for the domestic publications. This trend is more pronounced for the top-10 regions.

For the EU15 and the USA no separate values of the SI for international and domestic co-publications are available in Luwel (2015). To benchmark the PRC figure 1b shows the average values of the SI of the two subsets together for the top-10 regions and the 31 regions and these data for the intra-EU15 co-publications and for the USA the co-publications with at least two different states.

There is a strong contrast between the USA, EU15 and the PRC. For the USA the SI average value is in 2000 already substantially higher and the indicator has a stronger growth rate during the period. Between 2000 and 2005 the SI average values for the EU15 are the same as for the top-10 regions but from 2005 onwards the increase for the EU15 is much more pronounced. For the 31 regions together during most of the period the SI average values are stable, increasing slightly only during the most recent years.

To analyze the effect of the distance between regions on their propensity to collaborate the correlation between the co-publication matrix and the SI on the hand and the matrix of the distances between the regions is calculated.

For the two subsets the co-publication matrix and the distance matrix correlate at 1 % significance level for all years except for 2001. The same result is obtained for the SI. However for the top-10 regions the Mantel test shows that the co-publication matrix and the distance matrix as well as the SI and the distance matrix are not correlated even at 5% significance level.

Each region has its own international co-publication profile which can evolve over time. In table 2 for each of the 4 regions with the most international co-publications in 2015 the top 5 countries are listed. The USA is on top and around 50% of these regions' international co-publications has at least one address from this country in the byline. However the growth rate seems to level off, especially for GuangDong. A second observation is the decrease of the fraction of the publications in collaboration with researchers working in German institutes and the inverse tendency for Australia and the UK. Another striking observation is the absence of Japan among the top 5 except for co-publications with Shanghai but its share is decreasing rapidly. The same pattern is observed for most of the top regions as is illustrated in figure 2. For the other 6 regions in the top-10 the fraction of the international co-publications of the leading countries (except the USA) is given (publication year 2015). Between 44% and 47% of these regions' international co-publications are in collaboration with this country. Australia and the UK are competing for the second place. Only for Liaoning is Japan an important partner co-singing more than 10% of its international co-publications.

Discussion

The regional distribution of PRC's publications in journals processed for the WoS is highly skewed; the top-10 regions produce more than 75% of the total output. Although the growth rates are higher for the less productive regions, the share of the top-10 regions in the total regional publication output is only slowly reducing over the last 15 years. Similar skewness in the distribution of publications has been observed within the USA and between EU member states (Luwel, 2015).

The overall growth rate of both international co-publications and domestic publications slows down; comparing 2010 and 2015 this trend is more pronounced for international co-publications than for domestic publications. Their ratio is 31% in 2015. But again there is a large deviation between the top-10 regions with 33% and the third tier with 22%.

Using co-publications as a proxy in contrast with the USA and the EU the scientific collaboration between the regions is not increasing substantially. Only on international co-publications the PRC's regions are collaborating somewhat more intensely during the last couple of years. In the paper the evolution of the average value of the SI is presented; using the median value of this index and other similarity measures similar results are obtained.

For the 31 regions together the scientific collaboration measured by co-publications and geographical separation correlates very well. This result is strongly influenced by the regions with the lowest number of (co-)publications. The collaboration among the top-10 regions evolves independently from their geographical location; over the period 2000-2015 the co-publications matrix and SI are not correlated with the distance matrix.

The USA is the international partner par excellence but there are differences in regions' international co-publication profiles with the propensity to collaborate with Germany and Japan decreasing and a growing influence of Australia and the UK. Especially for Germany's and

Japan's decline no obvious explanation can be given. Australia with it is strong international higher education sector and neighboring Asian countries become for Chinese researchers more attractive as scientific partners.

To obtain a better understanding of the relationship between the geography and the scientific collaboration among the regions work is in progress to construct Mantel correlograms (Diniz-Filho, 2013); although they assume an underlying distribution gravity models could also be used (Scherngell, 2011). To benchmark in more detail the trends in PRC's co-publications the data for the EU and the USA will be broken down in international and domestic co-publications. Finally the analysis presented in this paper can also be made using the 10% most cited papers or for individual scientific disciplines.

References

Andersson D.E., Gunessee S., Matthiessen C.W. & Find S. (2014). The Geography of Chinese Science. *Environment and Planning A*, 46, 2950-2971.

Bellini N. & Hilpert U. (Ed.). (2013). *Europe's Changing Geography: The impact of Inter-regional Networks*. Abingdon: Routledge.

Diniz-Filho J.A.F., Soares T.N., Lima J.S., Dobrovolski R., Landeiro V.L., Pires de Campos Telles M., Rangel T.F. & Bini L.M. (2013). Mantel Test in Population Genetics. *Genetics and Molecular Biology*, 36, 475-485.

Glänzel W., Debackere K. & Meyer M. (2008). '*Triad ' or 'Tetrad'?* On Global Changes in a Dynamic Wolds. *Scientometrics*, 74, 71-88.

Haustein S., Tunger D., Heinrichs G. & Baelz G. (2011). Reasons for and Developments in International Scientific Collaboration: Does an Asia-Pacific Research Area Exits from a Bibliometric Point of View? *Scientometrics*, 86, 727-746.

Liang L. & Zhu L. (2002). Major Factors Affecting China's Inter-regional Research Collaboration: Regional Scientific Productivity and Geographical Proximity. *Scientometrics*, 55, 287-316.

Luukkonen T., Tijssen R.J.W., Persson O. & Sivertsen G. (1993). The Measurement of International Scientific Collaborations. *Scientometrics*, 28, 15-36.

Luwel M & van Wijk E. (2015), Scientific Europe Coming Together? *20th International Conference on Science and Technology Indicators, Lugano, Switzerland, September 2-4, 2015.*

National Science Board (2014). *Science and Engineering Indicators 2014*. Arlington VA: National Science Foundation (NSB 14-01).

Mantel N. (1967). The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, 27, 209-220.

Scherngell T. & Hu Y. (2011). Collaborative Knowledge Production in China: Regional Evidence from a Gravity Model Approach. *Regional Studies*, 45, 755-772.

Sun Y. & Cao C. (2015). Intra- and Inter-Regional Research Collaboration across Organizational Boundaries: Evolving Patterns in China. *Technological Forecasting and Social Change*, 96, 215-231.

Tang L. & Shapira P. (2011). *China–US scientific collaboration in nanotechnology: patterns and dynamics*. Scientometrics, 88, 1-16.

Van Eck, N.J., & Waltman, L. (2009). How to Normalize Cooccurrence Data? An Analysis of Some Well-Known Similarity Measures. *Journal of the American Society for Information Science and Technology*, 60, 1635-1651.

Wagner C.S., Bornmann L. & Leydesdorff L. (2015). Recent Developments on China-US Cooperation in Science. *Minerva*, 53, 199-214.

Waltman L., Tijssen R.J.W. & van Eck N.J. (2011). Globalisation of Science in Kilometers, *Journal of Informetrics*, 5, 574-582.

Table 1a. For the 31 regions, the number of domestic publications for the publication year 2000, 2005, 2010 and 2015 (full counting scheme). The sixth colon (%2015) gives each region's percentage of the total number of publications in 2015. The last three colons give the Percent (Straight-Line) Growth Rates for 2005-2000 (Gr 05-00), 2010-2005 (Gr 10-05) and 2015-2010 (Gr 15-10)

No	Region	2000	2005	2010	2015	% 2015	Gr 05-00	Gr 10-05	Gr 15-10
1	Prch-Beijing	6614	14925	24424	41069	17,95	125,7	63,6	68,2
2	Prch-Jiangsu	1687	4641	10678	21997	9,62	175,1	130,1	106,0
3	Prch-Shanghai	2989	7226	12737	20602	9,01	141,8	76,3	61,7
4	Prch-GuangDong	887	2487	6650	13421	5,87	180,4	167,4	101,8
5	Prch-Shandong	860	2873	5994	11840	5,18	234,1	108,6	97,5
6	Prch-Zhejiang	858	3601	6622	11316	4,95	319,7	83,9	70,9
7	Prch-Hubei	1080	3300	5963	11201	4,90	205,6	80,7	87,8
8	Prch-Shaanxi	820	2480	5481	10874	4,75	202,4	121,0	98,4
9	Prch-Sichuan	682	1929	4924	9442	4,13	182,8	155,3	91,8
10	Prch-Liaoning	948	2826	5081	8353	3,65	198,1	79,8	64,4
	Total 1-10	*17425*	*46288*	*88554*	*160115*	*70,00*	*165,6*	*91,3*	*80,8*
11	Prch-Tianjin	726	2198	3483	7285	3,18	202,8	58,5	109,2
12	Prch-Hunan	654	1873	3780	7069	3,09	186,4	101,8	87,0
13	Prch-Anhui	1101	2471	3587	6480	2,83	124,4	45,2	80,7
14	Prch-Jilin	875	2127	3550	6097	2,67	143,1	66,9	71,7
15	Prch-Heilongjiang	308	1408	3260	5765	2,52	357,1	131,5	76,8
16	Prch-Henan	275	963	2634	5633	2,46	250,2	173,5	113,9
17	Prch-Chongqing	185	761	2178	4592	2,01	311,4	186,2	110,8
18	Prch-Fujian	491	1304	2405	4579	2,00	165,6	84,4	90,4
19	Prch-Gansu	707	1410	2455	3680	1,61	99,4	74,1	49,9
20	Prch-Hebei	273	803	1601	3339	1,46	194,1	99,4	108,6
	Total 11-20	*5595*	*15318*	*28933*	*54519*	*23,83*	*173,8*	*88,9*	*88,4*
21	Prch-Jiangxi	87	417	1225	2638	1,15	379,3	193,8	115,3
22	Prch-Shanxi	179	667	1176	2365	1,03	272,6	76,3	101,1
23	Prch-Yunnan	277	575	1382	2271	0,99	107,6	140,3	64,3
24	Prch-Guangxi	104	346	977	2091	0,91	232,7	182,4	114,0
25	Prch-Xinjiang	74	189	661	1439	0,63	155,4	249,7	117,7
26	Prch-Guizhou	66	143	424	1043	0,46	116,7	196,5	146,0
27	Prch-Inner Mongolia	38	146	402	934	0,41	284,2	175,3	132,3
28	Prch-Hainan	22	67	232	624	0,27	204,5	246,3	169,0
29	Prch-Ningxia	12	43	110	335	0,15	258,3	155,8	204,5
30	Prch-Quinghai	16	78	157	307	0,13	387,5	101,3	95,5
31	Prch-Tibet	3	3	39	67	0,03	0,0	1200,0	71,8
	Total 21-31	*878*	*2674*	*6785*	*14114*	*6,17*	*204,6*	*153,7*	*108,0*
	Total	***23898***	***64280***	***124272***	***228748***	***100,00***	***169,0***	***93,3***	***84,1***

Table 1b. For the 31 regions, the number of international co-publications. The table has the same structure as table 1a

No	Region	2000	2005	2010	2015	% 2015	Gr 05-00	Gr 10-05	Gr 15-10
1	Prch-Beijing	1967	4103	8476	15673	22,30	108,6	106,6	84,9
2	Prch-Shanghai	687	1680	4072	8135	11,57	144,5	142,4	99,8
3	Prch-Jiangsu	468	1106	2654	6783	9,65	136,3	140,0	155,6
4	Prch-GuangDong	172	582	1799	4339	6,17	238,4	209,1	141,2
5	Prch-Hubei	257	704	1803	4064	5,78	173,9	156,1	125,4
6	Prch-Zhejiang	233	704	1837	3689	5,25	202,1	160,9	100,8
7	Prch-Shaanxi	170	453	1381	3103	4,42	166,5	204,9	124,7
8	Prch-Shandong	152	509	1259	2565	3,65	234,9	147,3	103,7
9	Prch-Sichuan	148	329	1150	2539	3,61	122,3	249,5	120,8
10	Prch-Liaoning	252	672	1245	2144	3,05	166,7	85,3	72,2
	Total 1-10	*4506*	*10842*	*25676*	*53034*	*75,46*	*140,6*	*136,8*	*106,6*
11	Prch-Anhui	246	499	1005	2054	2,92	102,8	101,4	104,4
12	Prch-Tianjin	140	430	872	1898	2,70	207,1	102,8	117,7
13	Prch-Hunan	74	308	804	1874	2,67	316,2	161,0	133,1
14	Prch-Heilongjiang	80	297	784	1595	2,27	271,3	164,0	103,4
15	Prch-Jilin	156	369	831	1422	2,02	136,5	125,2	71,1
16	Prch-Fujian	100	241	547	1404	2,00	141,0	127,0	156,7
17	Prch-Chongqing	27	119	444	1201	1,71	340,7	273,1	170,5
18	Prch-Henan	37	106	321	1042	1,48	186,5	202,8	224,6
19	Prch-Gansu	109	232	608	878	1,25	112,8	162,1	44,4
20	Prch-Yunnan	68	182	408	767	1,09	167,6	124,2	88,0
	Total 11-20	*1037*	*2783*	*6624*	*14135*	*20,11*	*168,4*	*138,0*	*113,4*
21	Prch-Hebei	33	116	269	556	0,79	251,5	131,9	106,7
22	Prch-Shanxi	36	117	163	556	0,79	225,0	39,3	241,1
23	Prch-Jiangxi	11	57	177	499	0,71	418,2	210,5	181,9
24	Prch-Guangxi	25	89	195	432	0,61	256,0	119,1	121,5
25	Prch-Xinjiang	16	39	139	321	0,46	143,8	256,4	130,9
26	Prch-Guizhou	24	44	111	280	0,40	83,3	152,3	152,3
27	Prch-Inner Mongolia	17	24	102	154	0,22	41,2	325,0	51,0
28	Prch-Hainan	2	8	44	144	0,20	300,0	450,0	227,3
29	Prch-Quinghai	2	20	40	88	0,13	900,0	100,0	120,0
30	Prch-Ningxia	1	10	30	60	0,09	900,0	200,0	100,0
31	Prch-Tibet	4	3	15	23	0,03	-25,0	400,0	53,3
	Total 21-31	*171*	*527*	*1285*	*3113*	*4,43*	*208,2*	*143,8*	*142,3*
	Total	***5714***	***14152***	***33585***	***70282***	***100,00***	***147,7***	***137,3***	***109,3***

Table 2. For the first 4 regions in the top-10 the 5 countries with the highest number of co-publications are listed as well as the ratio of this number and the region's total number of international co-publications for each uneven year between 2000 and 2015

Region	Country	2001	2003	2005	2007	2009	2011	2013	2015
Beijing	USA	0,40	0,43	0,47	0,46	0,46	0,48	0,51	0,52
Beijing	UK	0,09	0,09	0,12	0,10	0,10	0,10	0,12	0,12
Beijing	Germany	0,13	0,13	0,12	0,10	0,09	0,10	0,11	0,10
Beijing	Australia	0,04	0,05	0,05	0,07	0,06	0,07	0,08	0,09
Beijing	Canada	0,05	0,05	0,08	0,07	0,08	0,07	0,07	0,08
Jiangsu	USA	0,35	0,33	0,38	0,38	0,45	0,45	0,46	0,48
Jiangsu	Australia	0,05	0,06	0,06	0,07	0,06	0,09	0,09	0,09
Jiangsu	UK	0,06	0,07	0,06	0,08	0,09	0,09	0,08	0,09
Jiangsu	Canada	0,06	0,05	0,07	0,07	0,08	0,07	0,07	0,07
Jiangsu	Germany	0,06	0,10	0,08	0,08	0,08	0,07	0,07	0,07
Shanghai	USA	0,34	0,38	0,43	0,43	0,45	0,48	0,53	0,53
Shanghai	UK	0,08	0,08	0,09	0,08	0,10	0,09	0,10	0,11
Shanghai	Australia	0,06	0,06	0,03	0,05	0,06	0,06	0,08	0,09
Shanghai	Germany	0,12	0,10	0,08	0,08	0,08	0,08	0,08	0,09
Shanghai	Japan	0,24	0,20	0,18	0,15	0,12	0,10	0,09	0,09
GuangDong	USA	0,31	0,33	0,29	0,38	0,47	0,48	0,51	0,51
GuangDong	Australia	0,05	0,09	0,13	0,12	0,09	0,09	0,10	0,13
GuangDong	UK	0,07	0,07	0,08	0,08	0,08	0,09	0,08	0,11
GuangDong	Canada	0,06	0,08	0,09	0,08	0,06	0,07	0,07	0,08
GuangDong	Germany	0,19	0,11	0,08	0,07	0,05	0,07	0,07	0,08

Figure 1a. Evolution of the SI average value of the international co-publications (a) and the domestic co-publications (b) of the 31 regions and of the top-10 regions (c and d)

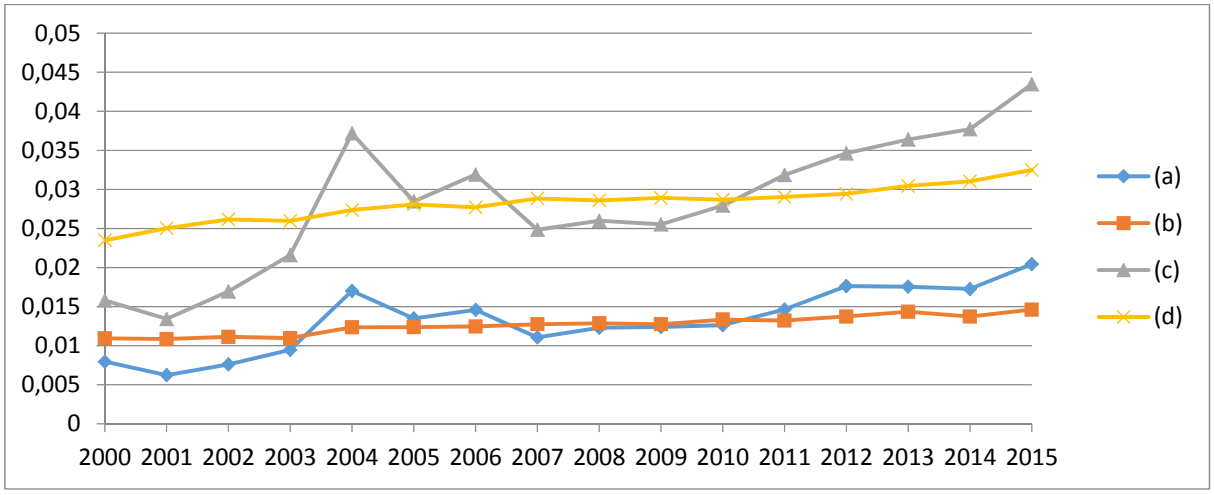

Figure 1b. Evolution of the SI average value of the co-publications of the 31 regions (a), the top-10 regions (b), the EU15 (c) and the USA (d)

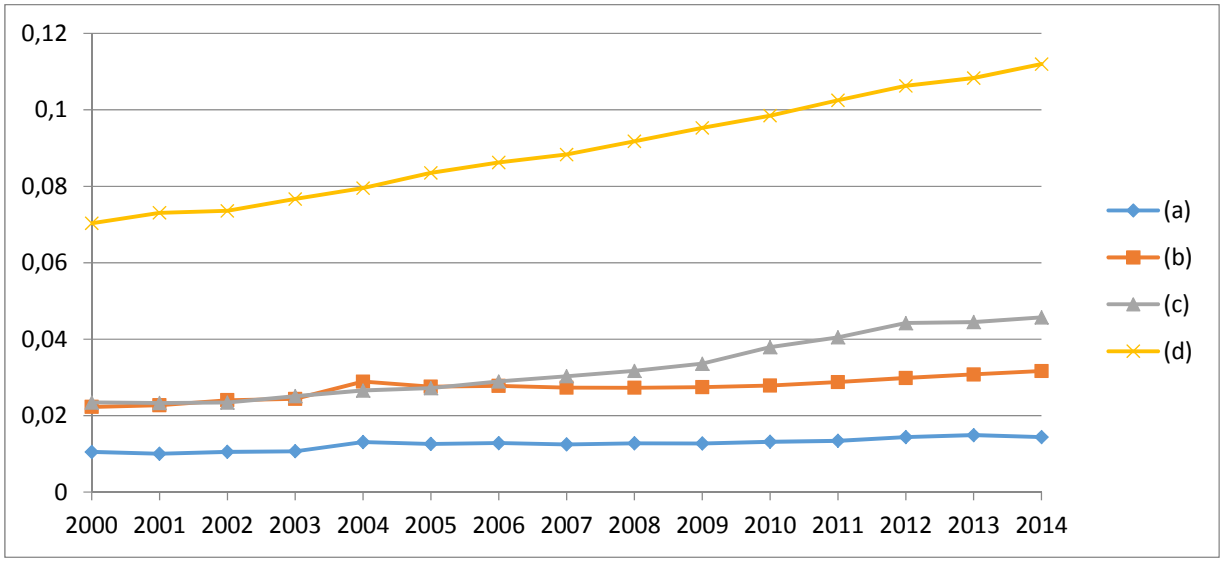

Figure 2. For Hubei, Liaoning, Shaanxi, Shandong, Sichuan and Zhejing the countries with the most co-publications in 2015 (after the USA, not shown in the graph) are given as well as the ratio of this number and the region's total number of international co-publications.