Collaboration benefits and international visibility of two-country-collaborations

Riechert, Mathias; Sirtes, Daniel; Lietz, Haiko

Veröffentlichungsversion / Published Version
Sammelwerksbeitrag / collection article

Collaboration Benefits and International Visibility of Two-Country-Collaborations[1]

Mathias Riechert — riechert@forschungsinfo.de / iFQ – Institute for Research Information and Quality Assurance, Schützenstr. 6a, D-10177 Berlin (Germany)

Daniel Sirtes — sirtes@forschungsinfo.de / iFQ – Institute for Research Information and Quality Assurance, Schützenstr. 6a, D-10177 Berlin (Germany)

Haiko Lietz — haiko.lietz@gesis.org / GESIS – Leibniz Institute for the Social Sciences, Sachsenhausen 6–8, D-50667 Köln (Germany)

Abstract

We introduce two novel ways to capture the impact benefits two countries receive from collaborations. For both indicators we compare the value of a specific collaboration with the value of average collaborations for each of the countries. As we restrict our analysis to only two-country collaborations and calculate the values for each scientific field individually, many of the problems introduced by former attempts of collaboration indicators dissipate. Additional to a mean based indicator (*Citation Benefit*) we introduce an *International Citation Share* indicator that measures the share of international citations on an item basis. By aggregating and correlating these indicators we show that two different factors of collaboration return, highly cited publications and a general domestic/international bias, *i.e.* the tendency of a publication to be cited more in the originating country, can be measured exclusively by these two indicators. These approaches open up the field to a new kind of deep analysis of scientific collaborations.

Introduction

Many studies imply that collaboration increases the amount of received citations (Chinchilla-Rodríguez, Vargas-Quesada, Hassan-Montero, González-Molina, & Moya-Anegóna, 2010; Glänzel, 2001, 2002; Hsu & Huang, 2011; Katz & Hicks, 1997; Lewinson & Cunningham, 1991; Narin, Stevens, & Whitlow, 1991; Persson, Glänzel, & Danell, 2004). In order to analyze the collaboration link's successfulness, the Relative Citation Eminence (RCE) has been introduced (Glänzel & Schubert, 2001). RCE is computed by dividing the mean observed citation rate of publications co-authored by countries X and Y by the geometric mean of the mean observed citation rates of X and Y. However, the RCE is a symmetric value and does not show which country benefits more from the collaboration or not at all. It was shown that countries differ in their scientific collaboration work depending on their size, citation culture, productivity and geographical proximity (Ding, Foo, & Chowdhury, 1998; Glänzel, Schubert, & Czerwon, 1999; Glänzel, 2001; Katz &

1 Thanks to Jörg Neufeld (iFQ) for useful discussions and Almuth Lietz (iFQ) for help with the figures.

Hicks, 1997; Luukkonen, Tijssen, Persson, & Sivertsen, 1993; Pao, 1981; Singh, 2005; Zhao & Guan, 2011). An asymmetric citation behavior is therefore to be expected and focus of the present paper. To measure aspects of this inequality, several indicators have been proposed recently. The *Citation Rate Increment from the Collaborator* (CRIC) and the *Domestic Citation Rate Comparison when Collaborating* (DCRCC) measure if a collaboration between countries O and P yields more citations from O to P and O to O, respectively. The *Domestic Impact Rate Increment when Collaborating* (DIRIC) measures the increment in average domestic citations relative to the non-collaboration case (Lancho-Barrantes, Guerrero-Bote, & De Moya-Anegón, 2012). The *International Collaboration Gain in Normalized Citation* (ICGNC) is the difference between the field normalized citation rate of collaborative and non-collaborative publications (Guerrero Bote, Olmeda-Gómez, & De Moya-Anegón, 2013).

We contend that there are several methodological problems with these indicators. CRIC, DCRCC, and DIRIC lack field normalization which skews the results toward the highly cited disciplines. Third collaborating countries are not controlled for. The focus, in the case of CRIC and DCRCC, on raw citation scores is very sensitive to extreme values. OCGNC is field normalized but its construction as a difference can be challenged. Finally, all indicators compare collaboration with non-collaboration. This may introduce a bias between international collaborating authors vs. non-collaborating ones.

We introduce methods where specific collaboration is compared with general collaboration. In the indicators to be described, the benefit for one country to collaborate with another is compared to the average[i] of its collaborations. We also avoid some of the methodological problems of the indicators described above. All our indicators are field normalized. Our set of publications includes only those with two collaborating countries. Additionally, in order to measure international visibility we introduce a citation share indicator which is not sensitive to extreme values. The comparison between the values of our indicators helps clarify the latent factors determining high collaboration returns and high international visibility.

Data and Methods

The publications for the analysis are drawn from the Competence Centre for Bibliometrics for the German Science System's bibliometric database based on Web of Science (WoS), published between 2005 and 2009 in journal articles or reviews in all fields. The focus of this study are publications which were collaboratively authored by exactly two countries, as multilateral collaboration requires a differentiated approach (Glänzel & De Lange 1997). 843,666 distinct publications from WoS fulfilled the above criteria. This selection does not exclude the authors with more than one affiliation as was shown problematic by Katz and Martin (1997). However, for our study this only becomes virulent if one author is affiliated with two countries and no other author has a different country affiliation. Our analysis shows that only 0.7 percent of authors have this kind of double country-affiliation, which makes this problem rather negligible.

A sliding citation window of three years was used. All citations were counted without self-citations to prevent self-citing effects from skewing the results (especially on international visibility). To reduce the effect of extreme values, only papers were selected that were in a set of at least 20 papers in a country-field and a country-country combination. The Publications below the threshold of 20 were discarded for both countries, but only for the respective field (as it can be present in different fields simultaneously). From the initial sample, 807,535 (95.7%) publications from 116 distinct countries and 222 distinct fields (using the subject classification scheme of WoS) were consequently used as the basis for the calculations.

The following five indicators are proposed:

Starting point for the *Collaboration Benefit (CB)* indicator construction are the citations per paper, which is associated with a field *f* and was penned by collaborators from countries *o* and *p*. This is the Mean Observed Citation Rate $MOCR_{opf}$. The $MOCR_{opf}$ is divided by the Mean Observed Citation Rate of country *o*'s co-authored papers in that field f. This allows for comparisons between different fields. The indicator for collaboration citation benefit of a country with another country in a field is therefore proposed as follows:

$$CB_{opf} = \frac{MOCR_{opf}}{MOCR_{of}^{co}}$$

$MOCR_{opf}$ Mean Observed Citation Rate of country *o*'s papers co-authored with country *p* in field *f*.

$MOCR_{of}^{co}$ Mean Observed Citation Rate of country *o*'s co-authored papers in field *f*.

To illustrate the construction of the indicator, the collaboration between the USA (*o*) and France (*p*) in the field of thermodynamics (*f*) is used: The USA and France have collaborated in 57 papers and received 301 citations (without self-citations) for them. Consequently, the $MOCR_{opf}$ is 5.28 for both countries in this field. This value is to be compared with the average citations the respective country receives, while collaborating with any one other country, in that field: While the USA has published 1,185 collaborative papers receiving 7,359 citations in the field thermodynamics ($MOCR_{USA}^{co} = 5.03$), France published 479 such papers receiving 2,407 citations ($MOCR_{FRA}^{co} = 6.24$). Therefore the Collaboration Benefit for the USA in collaboration with France in the field Thermodynamics is $CB_{USA,FRA,f} = 5.28/5.03 = 0.94$. while the corresponding Collaboration Benefit is $CB_{FRA,USA,f} = 5.28/6.24 = 0.89$. Both countries have fewer citations per paper in collaboration compared to the citations both countries generally receive in collaborated papers in the field of thermodynamics.

International Collaboration Benefit (CB_INT)

The second indicator we propose is similarly designed, but targets the international citations of the collaborated:

$$CB\_INT_{opf} = \frac{MOCR\_INT_{opf}}{MOCR\_INT_{of}^{co}}$$

$MOCR\_INT_{opf}$ International Mean Observed Citation Rate of object *o*'s papers co-authored with object *p* in field *f*.

$MOCR\_INT_{of}^{co}$ International Mean Observed Citation Rate of object *o*'s co-authored papers in field *f*.

For illustration the same example is used: The Collaboration Benefit for the USA in collaboration with France in the field Thermodynamics is $CB\_INT_{USA,FRA,f} = 3.63/4.48 = 0.85$ while the corresponding International Collaboration Benefit is $CB\_INT_{FRA,USA,f} = 4.70/4.38 = 1.05$. The USA has fewer international citations per paper in collaboration compared to the international citations normally received from collaborated papers in the field of thermodynamics. In contrast, France has more international citations compared to all international citations from collaborations. In terms of international citations France benefits more from the collaborations.

Domestic Collaboration Benefit (CB_DOM): The third proposed indicator is designed analogically to the International Collaboration Benefit to measure the Domestic Citation Benefit CB_DOM_{opf}.

$$CB\_DOM_{opf} = \frac{MOCR\_DOM_{opf}}{MOCR\_DOM_{of}^{co}}$$

$MOCR\_DOM_{opf}$ Domestic Mean Observed Citation Rate of country *o*'s papers co-authored with country *p* in field *f*

$MOCR\_DOM_{of}^{co}$ Domestic Mean Observed Citation Rate of country *o*'s co-authored papers in field *f*

International Citation Share (CS_INT):The fourth indicator we propose measures the Mean International Citation Share (MICS) of two collaborating countries in a field *f*. The main reason to use the share of international and domestic citations received in collaboration is that this is not dependent on highly cited papers as the share of international vs. domestic citations is weighed for each publication equally, disregarding how many citations an individual paper has yielded. In contrast to the *CB*, the *CS* therefore measures the typical international visibility of the collaboration. The *Item's International Citation Share* is computed by dividing the international citations per paper by all citations. The *Mean International Citation Share* is the average of the IICS in the field. The *International Citation Share* is consequently computed by dividing the *Mean International Citation Share* of the countries co-authored papers in the field *f* by the *Mean International Citation Share* of all publications of the country *o* in field *f*.

Item's International Citation Share (per paper *i*):

$$IICS_i = \frac{non - domestic\ citations_i}{all\ citations_i}$$

International Citation Share:

$$CS\_INT_{opf} = \frac{MICS_{opf}}{MICS_{of}^{co}}$$

$MICS_{opf}$ Mean International Citation Share of country *o*'s papers co-authored with country *p* in field *f*

$MICS_{of}^{co}$ Mean International Citation Share of country *o*'s co-authored papers in field *f*

Example: The USA and France did collaborate in 51 papers. This number differs from the one above as only cited papers are counted for citations shares. In the field Thermodynamics these papers received 301 total citations. For each paper the IICS is computed and subsequently averaged over all papers resulting in a $MICS_{USA,FRA,f} = 0.68$ and a $MICS_{FRA,USA,f} = 0.90$. The average value for the USA in the field is $MICS_{USA,f}^{co} = 0.68$ and for France $MICS_{FRA,f}^{co} = 0.87$. The resulting *CS_INT* values are $CS\_INT_{FRA,USA,f} = 1.00$ and $CS\_INT_{FRA,USA,f} = 1.03$.

Domestic Citation Share (CS_DOM)

The Domestic Citation Share indicator is constructed analogously to *CS_INT* only with the Item's *Domestic* Citation Share as basis.

Aggregation: The proposed indicators are calculated for each collaboration relationship in each field. For each collaboration pair, the number of publications *n* is used for a weighted average, *i.e.* a field-normalized aggregation. Therefore, each indicator can subsequently be aggregated to the field-independent country-country level. Finally, again a weighted aggregation on the collaborating country p is possible in order to get indicators on the country level. It is possible to aggregate these indicators to the country-field and field level as well. As the intention of our study is to analyze the collaboration for countries and not for fields we focus on this aggregation path in the following.

Results

On the country-country-field level, 181,110 different combinations with the five described indicators were computed in 116 countries and 222 fields. Table 1 shows the descriptive statistics for the indicators on the country-country-field level. It is evident that, as to be expected, the CS indicators have less dispersion than the CB indicators.

Table 1. Indicator Descriptive Statistics

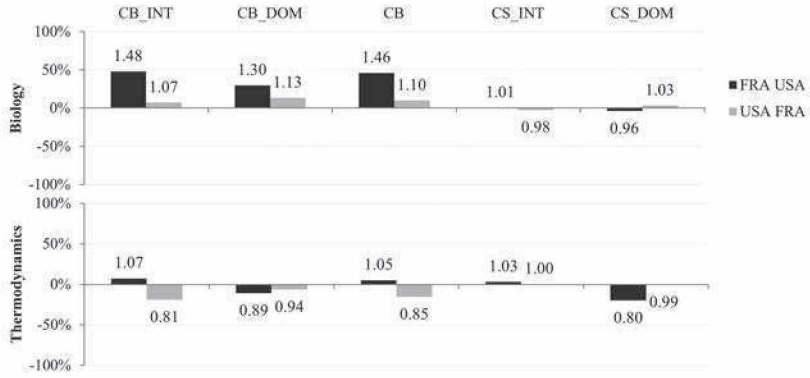
	N	CB_INT	CB_DOM	CB	CS_INT	CS_DOM
Min:	1	0	0	0	0	0
1st quartile:	1	0.3407	0	0.3503	0.975	0
Median:	3	0.707	0.4059	0.7112	1.028	0.571
Mean:	14.39	0.9006	0.8868	0.8975	1.005	0.98
3rd quartile:	10	1.1409	1.1288	1.1381	1.075	1.218
Max:	2346	106.8925	78.1844	103.3742	2.773	48

Table 2. Indicator Values in Different Fields:

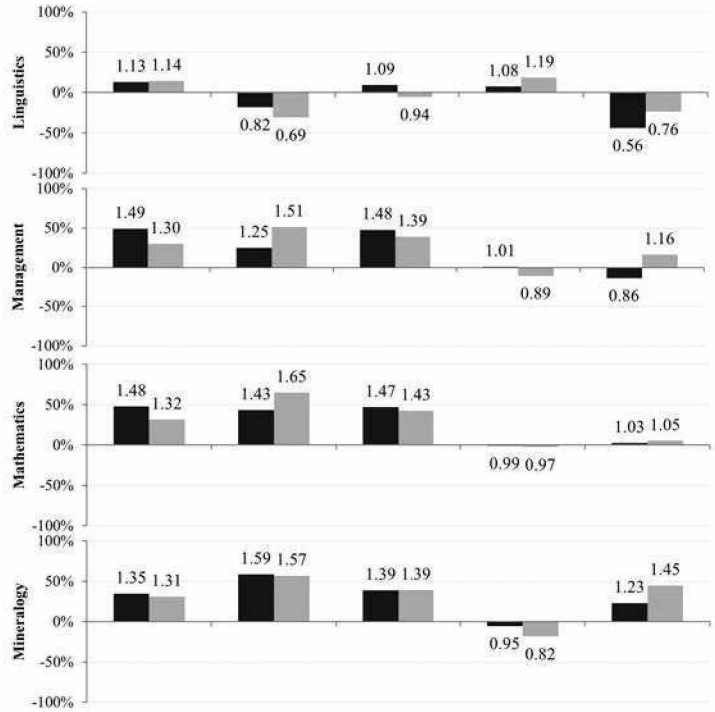
Country O	Country P	Field	CB_INT	CB_DOM	CB	CS_INT	CS_DOM
FRA	USA	Biology	1.48	1.30	1.46	1.01	0.96
USA	FRA	Biology	1.07	1.13	1.10	0.98	1.03
FRA	USA	Thermodynamics	1.07	0.89	1.05	1.03	0.80
USA	FRA	Thermodynamics	0.81	0.94	0.85	1.00	0.99
FRA	USA	Linguistics	1.13	0.82	1.09	1.08	0.56
USA	FRA	Linguistics	1.14	0.69	0.94	1.19	0.76
FRA	USA	Management	1.49	1.25	1.48	1.01	0.86
USA	FRA	Management	1.30	1.51	1.39	0.89	1.16
FRA	USA	Mathematics	1.48	1.43	1.47	0.99	1.03
USA	FRA	Mathematics	1.32	1.65	1.43	0.97	1.05
FRA	USA	Mineralogy	1.35	1.59	1.39	0.95	1.23
USA	FRA	Mineralogy	1.31	1.57	1.39	0.82	1.45

The results shown in Table 2 are on the already introduced example of the collaboration between the USA and France. Both countries collaborated in 205 disciplines. For reasons of space we will show only a limited subset of fields. The indicator values are depicted as a graph in Figure 1.

Figure 1. Indicator Values for the USA and France



... Continuation Figure 1



CB: France benefits in all six fields compared to its overall collaborations when collaborating with the USA, albeit the benefit varies considerably between the fields (1.48 in management vs. 1.05 in thermodynamics). The USA benefits less from the collaboration with France and has lower citation rates in the fields of thermodynamics and linguistics compared to their usual citations in the fields. The benefit in terms of international citation (*CB_INT*) is close to the *CB* in the fields analyzed, although there are some variations for the USA. The domestic citations are observably different for the fields. The collaboration in the field mathematics is, for example, characterized by high benefits for both countries 1.43 and 1.65 as well as in mineralogy while collaboration in linguistics and thermodynamics seems to attract fewer domestic citations.

CS: The results for the citation share indicator differ substantially from the CB results. In coherence with Table 1, the *CS_INT* and *CS_DOM* have less variance compared to the CB indicators. The *CS_INT* is only above 1 in the field of linguistics for the USA; therefore only in this field the USA receives a higher International Citation Share compared to their overall collaborations. Mineralogy is an example for how both indicator types can differ in the same field. Although both countries benefit from the collaboration in terms of international, domestic and overall citations, the *International Citation Share* is clearly lower compared to their overall collaborations. This conspicuous difference between the indicators is the subject of a further correlation analysis, which is conducted in order to obtain an interpretation of the indicators in the following.

Table 3 shows the Pearson correlation coefficient for the proposed indicators computed for all 181,110 combinations. The results show that none of the indicators is correlating with the number of publications (*N*). The highest correlation (0.991) is found between the *CB_INT* and *CB* indicator. There are considerably high correlations between *CS_INT* and *CS_DOM* (– 0.722) and between *CB_DOM* and *CS_DOM* (0.615). Additionally there are minor correlations between *CB_INT* and *CB_DOM* (0.331), *CB_DOM* and *CB* (0.414) and *CB_DOM* and *CS_INT* (– 0.417). All other indicator combination correlations are below 0.1.

Table 3. Indicator Correlation

	CB_INT	*CB_DOM*	*CB*	*CS_INT*	*CS_DOM*
N	0.026	0.018	0.027	-0.008	0.003
CB_INT		0.331	0.991	0.106	-0.071
CB_DOM			0.414	-0.417	0.615
CB				0.028	-0.016
CS					-0.722

Discussion

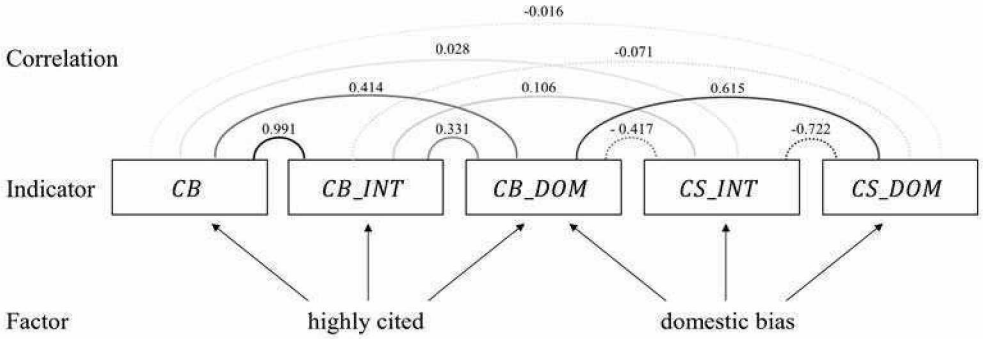
The results of the analyzed collaboration between the USA and France have shown a multifaceted collaboration. The high variation between the values in different fields reinforces the argument that collaboration on the basis of citations should be measured field-specific to prevent skewed results. An aggregation of citation-based collaboration indicators should therefore be based on the weighted field values.

While both countries observably do benefit in the fields of mineralogy, mathematics and management, France is benefiting substantially more in biology in terms of citations. The collaboration in the fields of linguistics and thermodynamics feature fewer citation benefits for both countries. In terms of domestic citation benefits, both countries have even fewer citations than their usual collaboration papers receive. An important difference is visible when comparing the citation benefits to the citation shares. As has been shown exemplarily, the share of citation per item differs substantially from the citations per paper. The collaborations of USA and France in the fields of mathematics and management show how the citations per paper can increase while the citation share is stable or even declining. We infer that the indicators measure different factors of collaboration, and have therefore to be analyzed in more detail.

Figure 2 depicts the correlations of the indicators. From the correlations between the citation benefit and the citation share indicators there are following points that can be learned. *CB* and *CB_INT* correlate very strongly. The non-domestic publication share, *i.e.* the share of all publications where the country at question is not involved is almost always the majority and in over 96%

of country-field combinations it amounts to over 90%. Therefore it is not surprising that both indicators will coincide.

Figure 2. Indicator Factor Dependencies



The citation benefit indicators are most influenced by very successful publications, as the mean is calculated over the set of papers in the combination. Therefore, a few extremely highly cited publications will skew the *CB* indicator to a very positive value. In contrast, the citation share indicator is not affected by these 'lucky few.' The share of international vs. domestic citations is weighed for each publication equally, disregarding how many citations an individual paper has yielded. Therefore the citation share indicator gives us a better picture of the *typical* distribution of citations rather than just reiterating that there are a couple of highly cited publications in the mix. The *International Citation Share* indicator being very lowly correlated with the *International Citation Benefit* indicator while quite highly negatively correlated with the domestic citation benefit indicator is a strong indication that the two international indicators truly measure two different latent dimensions.

The prior probability to be cited domestically is very low and therefore the *International Citation Benefit* is hardly affected by domestic citations. On the other hand the *Domestic Citation Benefit* is influenced by two factors: very highly cited papers will also have quite a few domestic ones, and for lowly cited papers it will show whether these few citations are over-proportionally domestic, this is what we call a *domestic bias*. The international citation share does only measure the second aspect as it does not measure highly cited papers more highly than lowly cited and therefore it really picks up on the domestic vs. international orientation of publications. This second, overlapping factor of Domestic Citation Benefit and International Citation Share is how the relatively high negative correlation between the two can be explained.

In conclusion, we suggest using only two of these indicators *CB* and *CS_INT* (or *CS_DOM*) in order to capture the two main factors that influence collaboration impact indicators, highly cited

publications and the general domestic bias. In this short paper we have only dipped into the vast possibilities of analyzing these indicators on different aggregation levels and as the basis of an in-depth *directed* network analysis (The manuscript of this analysis is in preparation).

References

Albarrán, P., Crespo, J. A., Ortuño, I., & Ruiz-Castillo, J. (2011). The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics*, 88(2), 385–397.

Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Hassan-Montero, Y., González-Molina, A., & Moya-Anegóna, F. (2010). New approach to the visualization of international scientific collaboration. *Information Visualization*, 9(4), 277–287.

Ding, Y., Foo, S., & Chowdhury, G. (1998). A bibliometric analysis of collaboration in the field of information retrieval. *International Information and Library Review*, 30, 367–376.

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1), 69–115.

Glänzel, W. (2002). Coauthorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. Katholieke Universiteit Leuven.

Glänzel, W. & De Lange, C. (1997). Modelling and measuring multilateral co-authorship in international scientific collaboration. Part II. A comparative study on the extent and change of international scientific collaboration links. *Scientometrics*, 40(3), 605–626.

Glänzel, W., & Schubert, A. (2001). Double effort = Double impact? A critical view at international co-authorship in chemistry. *Scientometrics*, 50(2), 199–214.

Glänzel, W., Schubert, A., & Czerwon, H. J. (1999). A bibliometric analysis of international scientific cooperation of the European Union (1985–1995). *Scientometrics*, 45(2), 185–202.

Guerrero Bote, V. P., Olmeda-Gómez, C., & De Moya-Anegón, F. (2013). Quantifying the benefits of international scientific collaboration. *Journal of the American Society for Information Science and Technology*, 64(2), 392–404.

Hsu, J., & Huang, D. (2011). Correlation between impact and collaboration. *Scientometrics*, 86(2), 317–324.

Katz, J. S., & Hicks, D. (1997). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, 40(3), 541–554.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research policy*, 26(1), 1–18.

Katz, J. S. (2012). Scale-Independent Measures: Theory and Practice. *Proceedings of the 17th International Conference on Science and Technology Indicators*,1–19.

Lancho-Barrantes, B. S., Guerrero-Bote, V. P., & De Moya-Anegón, F. (2012). Citation increments between collaborating countries. *Scientometrics*, 1–15.

Lewinson, G., & Cunningham, P. (1991). Bibliometric studies for the evaluation of trans-domestic research. *Scientometrics*, 21(2), 223–244.

Luukkonen, T., Tijssen, R. J. W., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28(1), 15–36.

Narin, F., Stevens, K., & Whitlow, E. S. (1991). Scientific co-operation in Europe and the citation of multinationally authored papers. *Scientometrics*, 21(3), 313–323.

Pao, M. L. (1981). Co-authorship as communication measure. *Library Research*, 2(4), 327–338.

Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421–432.

Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management science*, 51(5), 756–770.

Zhao, Q., & Guan, J. (2011). International collaboration of three 'giants' with the G7 countries in emerging nano-biopharmaceuticals. *Scientometrics*, 87(1), 159–170.

i Citation averages may happen to be power-law distributed (Albarrán, Crespo, Ortuño, & Ruiz-Castillo, 2011). Averages only give a valid characterization as long as the Central Limit Theorem holds (Katz, 2012). For the calculation and interpretation of the presented indicators one must be aware of a possible bias.

Erratum:

There has been a deplorable mix-up in the numeric examples:

p. 304:

To illustrate the construction of the indicator, the collaboration between the USA (*o*) and France (*p*) in the field of thermodynamics (*f*) is used: The USA and France have collaborated in 57 papers and received 301 citations (without self-citations) for them. Consequently, the $MOCR_{opf}$ is 5.28 for both countries in this field. This value is to be compared with the average citations the respective country receives, while collaborating with any one other country, in that field: While the USA has published 1,185 collaborative papers receiving 7,359 citations in the field thermodynamics ($MOCR^{co}_{USA} = 6.21$), France published 479 such papers receiving 2,407 citations ($MOCR^{co}_{FRA} = 5.03$). Therefore the Collaboration Benefit for the USA in collaboration with France in the field Thermodynamics is $CB_{USA,FRA,f} = 5.28/6.21 = 0.85$ while the corresponding Collaboration Benefit for France is $CB_{FRA,USA,f} = 5.28/5.03 = 1.05$. Thus, for France a collaboration with the USA is more beneficial than average, while the USA have fewer citations per paper in collaboration with France than in all collaborated papers in the field of thermodynamics.

p. 305

For illustration the same example is used: The Collaboration Benefit for the USA in collaboration with France in the field Thermodynamics is $CB\_INT_{USA,FRA,f} = 3.63/4.48 = 0.81$ while the corresponding International Collaboration Benefit is $CB\_INT_{FRA,USA,f} = 4.70/4.38 = 1.07$.