

Going Back in Time: Understanding Patterns of International Scientific Collaboration

Ronald Rousseau

University of Antwerp, Faculty of Social Sciences, B-2020 Antwerp, Belgium and KU Leuven, Facultair Onderzoekscentrum ECOOM, Naamsestraat, Leuven B - 3000, BELGIUM.

ABSTRACT

We study the interpretation of the observed/expected ratio as an indicator of international scientific collaboration. This indicator cannot be considered as representing the affinity for collaboration, in the sense of ‘the more the better’, between two countries.

Keywords: International collaboration, Observed/expected ratio, Indicators, Activity index, Axiomatic approach.

Correspondence

Ronald Rousseau

University of Antwerp, Faculty of Social Sciences, B-2020 Antwerp, Belgium and KU Leuven, Facultair Onderzoekscentrum ECOOM, Naamsestraat 61, Leuven B 3000, BELGIUM.
Email id: ronald.rousseau@uantwerpen.be

Received: 20-01-2021

Revised: 13-02-2021

Accepted: 21-02-2021

DOI: 10.5530/jsires.10.1.15

INTRODUCTION

In 1992 Luukkonen, Persson and Sivertsen published a well-known article entitled “Understanding patterns of international scientific collaboration” in the journal *Science, Technology, and Human Values*.^[1] Nowadays (January 2021) it is cited 326 times in the Web of Science (WoS). This places the article in the top position of articles published in this journal in 1992, and at the 11th place of all articles in this journal published since 1980.

Luukkonen *et al.*^[1] noted that international scientific collaboration has increased both in volume and importance, a point that holds even more nowadays. The authors wonder how one might explain country-to-country differences in the rates of international co-authorship and patterns of collaboration within the global network of collaborating countries. They draw the readers’ attention to cognitive, social, historical, geopolitical, and economic factors as potential determinants of the observed patterns. Indeed there is no single main factor but many factors that influence the propensity of countries to collaborate internationally. Luukkonen *et al.*^[1] offers an important tool for monitoring the effects of government science policy aiming to enhance international collaboration.

In their article, to study the importance of specific bilateral relations relative to all such relations within the whole network, the authors define the observed/expected ratio of co-authorship (*OER*) for a pair of countries (X,Y) in the set of publications under study as:

$$OER(X, Y) = \frac{\left(\frac{C_{X,Y}}{C_X} \right)}{\left(\frac{C_Y}{T} \right)} = \frac{\left(\frac{C_{X,Y}}{C_Y} \right)}{\left(\frac{C_X}{T} \right)} \quad (1)$$

where $C_{X,Y}$ denotes the number of collaborations between countries X and Y (whole counting); C_X the total number of collaborations country X has with all other countries, C_Y the total number of collaborations country Y has with all other countries, and T the total number of pairwise country collaborations in the set of publications under study. Hence T is the number of links in the weighted collaboration network, with countries as its nodes. It is assumed that C_X and C_Y are different from zero and for all countries X, $C_{X,X}$ is set equal to zero. Remark that in this counting scheme a publication with 5 authors, 2 from country X, 2 from country Y and 1 from country Z yields 3 country collaborations, namely X-Y, X-Z, and Y-Z. Clearly, *OER* is a relative indicator.

One may observe that formula (1) is one of the many variations of the activity (AI) or Balassa index,^[2-5] which we recall for completeness sake as formula (2):

$$AI = \frac{\left(\frac{O_{CD}}{O_D} \right)}{\left(\frac{O_C}{O_W} \right)} = \frac{\left(\frac{O_{CD}}{O_C} \right)}{\left(\frac{O_D}{O_W} \right)} \quad (2)$$

where:

O_{CD} denotes the number of publications by country C in domain D during a given publication window;

Copyright

© The Author(s). 2021 This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

O_D denotes the total number of publications in the world in domain D during the same publication window;

O_C denotes the number of publications – in all domains – by country C during the same publication window;

O_W denotes the total number of publications in the world and in all domains during this publication window.

However, to study the relative importance of country Y for country X while taking the other relations into account, the authors of^[1] used an asymmetric form of *OER*. We denote it as *AOER*. It is defined as:

$$AOER(X, Y) = \frac{\left(\frac{C_{X,Y}}{C_X}\right)}{\left(\frac{C_Y}{T - C_X}\right)} = \frac{\left(\frac{C_{X,Y}}{C_Y}\right)}{\left(\frac{C_X}{T - C_X}\right)} \quad (3)$$

Purely mathematically we have to point out that if country X is the only country with whom other countries collaborate ($T = C_X$), then *AOER* must be set equal to zero. The *AOER(X,Y)* indicator reflects the relative importance of country Y to country X, calculated as the share of country Y within all collaborations of country X divided by the share of country Y within all collaborations in the network (minus the collaborations of X). So, the size of the X-Y relation within X’s network is compared to the size of Y within the whole network, excluding X.

A PROBLEM WITH THE INTERPRETATION OF AOER

It was shown in^[4,6] that the activity index (AI) is not a good indicator for research policy applications. Its main problem is that, in the context of the research production of a given country in a given field, it depends on the activities of other countries in other fields (O_W).

AOER does not suffer from the same problems as the general AI, as the aspect of “other fields” does not play a role. Yet, it seems that this indicator has other problems in interpretation. Consider the following collaboration matrix, dealing with four countries:

	A	X	Y	B
A	0	2	0	14
X	2	0	2	0
Y	0	2	0	2
B	14	0	2	0

where $C_{X,Y} = 2$; $C_X = 4$; $C_Y = 4$ and $T = 20$. Then, we first we have that $AOER(X,Y) = (2/4)/(4/16) = 2$; and when adding two collaborations between X and Y *AOER(X,Y)* becomes: $(4/6)/(6/16) = 16/9 < 2$. So, two extra collaborations between

X and Y decrease the relative importance of country Y for country X, which seems contradictory. In the next section, we study this “anomaly” in more detail.

AOER IS NOT A MEASURE FOR THE ABSOLUTE NUMBER OF COLLABORATION LINKS OF COUNTRY Y WITH COUNTRY X

Intuitively, one might think, or maybe want, that a measure of collaborativeness between two countries should increase if these two countries increase their collaboration, as in the above example. This may or may not be a good property, but we prove that the *AOER* and the *OER* do not meet this requirement. Indeed, we shall show that from a certain natural number n on, where n denotes the number of new collaborations between countries X and Y, *AOER(X,Y)* decreases, with limit zero.

We compare $AOER(X, Y) = \frac{\left(\frac{C_{X,Y}}{C_X}\right)}{\left(\frac{C_Y}{T - C_X}\right)}$ with

$$AOER_n(X, Y) = \frac{\left(\frac{C_{X,Y} + n}{C_X + n}\right)}{\left(\frac{C_Y + n}{(T + n) - (C_X + n)}\right)}, \quad \text{where country}$$

X and Y have n more collaborations while all other data stay the same. We wonder for which natural number n $AOER(X, Y) > AOER_n(X, Y)$, i.e., when do extra collaborations between X and Y lead to a decrease of *AOER(X,Y)*?

For simplicity, we write $C_{X,Y}$ as a > 0, C_X as b, and C_Y as c. Then we have to solve:

$$\frac{a}{b \cdot c} > \frac{(a+n)}{(b+n) \cdot (c+n)}$$

This becomes: $a \cdot (b+n) \cdot (c+n) > b \cdot c \cdot (a+n)$,
 or $abn + can + an^2 > bcn$
 $an > bc - ab - ac$

And finally: $n > \frac{bc - a(b+c)}{a}$

Clearly, if $bc < a \cdot (b+c)$ this already happens for $n = 1$. We consider two examples. Let $a = 2$, $b = 3$, $c = 4$ and $T = 100$. Then $bc = 12 < a \cdot (b+c) = 14$. $AOER(X,Y) = 16.67$ and $AOER_1(X,Y) = 15.00$. More generally, let $a = 1$, $b = 3$, $c = 4$ and T any number larger than or equal to 4 (we take $T = 100$). Then, after an initial increase, a decrease in *AOER* (with respect to the initial value) happens for $n > 5$. Table 1 shows the corresponding values with $n = 0$ as the starting value. We

Table 1: An example showing the influence of increased collaboration between countries X and Y on $AOER(X,Y)$.

n	a	b	c	T	$AOER_n(X,Y)$
0	1	3	4	100	8.33
1	2	4	5	101	10.00
2	3	5	6	102	10.00
3	4	6	7	103	9.52
4	5	7	8	104	8.93
5	6	8	9	105	8.33
6	7	9	10	106	7.78
7	8	10	11	107	7.27
10	11	13	14	110	6.04
100	101	103	104	200	0.94
500	501	502	504	600	0.20

see that $AOER_5(X,Y)$ is still equal to $AOER(X,Y)$, but from $n = 6$ on, $AOER_n(X,Y) < AOER(X,Y)$.

As $\lim_{n \rightarrow \infty} \frac{(a+n).T}{(b+n).(c+n)} = 0$, $AOER_n(X, Y)$ tends to zero.

We observe that for $OER(X,Y)$ $\lim_{n \rightarrow \infty} \frac{(a+n).(T+n)}{(b+n).(c+n)} = 1$. So if

the collaboration between countries X and Y increases more and more their OER value will tend to 1. Depending on the original value of OER this would imply either an increase or a decrease.

OTHER PROPERTIES OF $AOER(X,Y)$

In this section, we present some properties of $AOER(X,Y)$ but warn the reader that this indicator is not a measure for the affinity (here understood as absolute increase in numbers of collaborated articles) between two countries. Recall that $a \leq b \leq T$ and $a \leq c \leq T$. To avoid uninteresting cases we moreover assume that $T > \max(b,c)$.

a) If b increases (by an amount of $p > 0$) at the cost of collaborations between others (hence T stays the same), and also a stays the same then $AOER(X,Y)$ decreases.

Proof. We have to check if $\frac{a.(T-b)}{b.c} > \frac{a.(T-(b+p))}{(b+p).c}$.

This inequality is equivalent with $(T-b).(b+p) > (T-b-p).b$

or: $Tb+Tp-b^2-bp > Tb -b^2 - bp$

or $Tp > 0$, which is clearly correct.

b) If b increases (by an amount of $p > 0$) and all other collaborations stay the same (hence T becomes $T+p$), and also a stays the same then $AOER(X,Y)$ decreases.

Proof. We have to check if $\frac{a.(T-b)}{b.c} > \frac{a.((T+p)-(b+p))}{(b+p).c}$.

or: $\frac{1}{b} > \frac{1}{(b+p)}$, which is clearly correct.

c) Assume that only T increases (by an amount of $p > 0$), i.e. some extra links are introduced not involving X or Y. Then $AOER(X,Y)$ increases.

Proof. We have to check if $\frac{a.(T-b)}{b.c} < \frac{a.(T+p)-b}{b.c}$, which is obviously correct.

Yet we warn that the following property d does not necessarily hold.

d) Assume that country Y increases its links (by an amount of $p > 0$) with other (not X) countries, at the cost of links with country X (so c and T stay the same). Hence a becomes $a-p$ (of course $a \geq p$), and b becomes $b-p$. Does this lead to a decrease in the value of $AOER(X,Y)$, because a has decreased? The answer is that this is not necessarily the case.

We have to compare $\frac{a.(T-b)}{b.c}$ and $\frac{(a-p).(T-(b-p))}{(b-p).c}$. For

$a = 100, b = 120, p = 30, T = 420$, (and $c = 300$) the left-hand side is: 0.833, while the right-hand side is 0.856 which is larger.

DISCUSSION AND CONCLUSION

1) Can one use or invent other measures instead of (1) and (3) for which an increased collaboration between two countries leads to an increase? That would, from a bilateral point of view, be more satisfactory. Could the F-measure, introduced in^[6] be used? Probably not. Or should one prefer Salton's or Jaccard's measure as suggested in Luukkonen *et al.*'s follow-up paper.^[7]

2) Could a similar study be done using fractional counting, instead of whole counting for country collaborations? Could^[8] be useful for this?

3) In^[7] the authors point out that in international collaboration studies bilateral as well as multilateral aspects are of interest. Moreover, one should certainly consider absolute and relative values.

We admit that we performed an ad hoc study. There is no required set of axioms, agreed upon by most colleagues, for an acceptable measure of relative research collaboration. We think that this is a general issue for informetric indicators. One should know the properties of indicators, leading to advantages and disadvantages for a given application. Vice versa, one should agree on a set of axioms for acceptable indicators of a certain type (for a certain purpose) and then find out which indicators are acceptable or invent new ones that are. In this context, we recall the case of the h-index. This index has been

characterized axiomatically, e.g., by Woeginger^[9] and a set of axioms, that excluded the h-index, for consistent rankings of departments has been proposed.^[10] Yet, even then, users of such indicators may decide that an indicator is “good enough” for practical use, say in a heuristic way (for a precise description of the term “heuristic” we refer the reader to ^[11]). Few indicators meet all possible requirements so that practitioners in our field often apply indicators that are PAC (probably approximately correct).^[12]

Concurring with the authors who constructed the *AOER* in 1992, we conclude that the observed/expected ratio as used in^[1] reflects the relative importance of country Y to country X, calculated as the share of country Y within all collaborations of country X divided by the share of country Y within all collaborations in the network (minus the collaborations of X). So, the size of the X-Y relation within X's network is compared to the size of Y within the whole network minus X. This indicator should not be considered as a measure of absolute affinity between countries X and Y.

ACKNOWLEDGEMENT

The author thanks Gunnar Sivertsen and Leo Egghe for helpful discussions about the content of this article.

CONFLICT OF INTEREST

The author declares no conflict of interest.

REFERENCES

1. Luukkonen T, Persson O, Sivertsen G. Understanding patterns of international scientific collaboration. *Science, Technology, Human Values*. 1992;17(1):101-26.
2. Balassa B. Trade liberalisation and 'revealed' comparative advantage. *Manchester School of Economic and Social Studies*. 1965;33(2):99-123.
3. Frame JD. Mainstream research in Latin America and the Caribbean. *Interciencia*. 1977;2(3):143-8.
4. Rousseau R, Yang LY. Reflections on the activity index and related indicators. *Journal of Informetrics*. 2012;6(3):413-21.
5. Rousseau R. Balassa = revealed comparative advantage = activity. *Scientometrics*. 2019;121(3):1835-6.
6. Rousseau R. The F-measure for research priority. *Journal of Data and Information Science*. 2018;3(1):1-18.
7. Luukkonen T, Tijssen RJW, Persson O, Sivertsen G. The measurement of international scientific collaboration. *Scientometrics*. 1993;28(1):15-36.
8. Rousseau R, Zhang L. Bilateral co-authorship indicators based on fractional counting. *Journal of Data and Information Science*. 2021. DOI: 10.2478/jdis-2021-0005.
9. Woeginger GJ. An axiomatic characterization of the Hirsch-index. *Mathematical Social Sciences*. 2008;56(2):224-32.
10. Bouyssou D, Marchant T. Ranking scientists and departments in a consistent manner. *Journal of the American Society for Information Science and Technology*. 2011;62(9):1761-9.
11. Bornmann L, Marewski JN. Heuristics as conceptual lens for understanding and studying the usage of bibliometrics in research evaluation. *Scientometrics*. 2019;120(2):419-59.
12. Rousseau R. Citation data as a proxy for quality or scientific influence are at best PAC (probably approximately correct). *Journal of the Association for Information Science and Technology*. 2016;67(12):3092-4.