



Using the weighted Lorenz curve to represent balance in collaborations: the BIC indicator

Ronald Rousseau^{1,2} · Lin Zhang^{1,3,4}  · Gunnar Sivertsen⁵

Accepted: 19 September 2022 / Published online: 10 October 2022
© Akadémiai Kiadó, Budapest, Hungary 2022

Abstract

This article aims to add to the body of indicators used to study collaboration in science. We propose a new indicator to measure balance in collaboration (BIC) which is based on the Gini evenness index for a weighted Lorenz curve. The new indicator of balance builds upon and extends the use of, our previously introduced indicator of relative intensity of collaboration (RIC). We present examples based on the collaboration network between the twenty largest countries contributing to science during 2000–2020.

Keywords International collaboration · Weighted Lorenz curve · Weighted Gini index · RIC-indicator

Introduction

The measurement and analysis of research collaboration by using co-authorship links in scientific papers have been practiced in bibliometric research for over 40 years. Aggregate statistics can be derived showing the development of collaboration patterns among institutions and countries within the global network. At these levels, indicators of research collaboration have proved useful for national policy analysis and organizational strategic development.

In Fuchs et al. (2021), we recently reviewed the literature on collaboration indicators and found reason to modify an often-used indicator that was first introduced in the article entitled “Understanding patterns of international scientific collaboration” by Luukkonen et al. (1992). In this article, the authors define the relative importance of country *Y* for country *X* as:

✉ Lin Zhang
zhanglin_1117@126.com

¹ Department MSI and Centre for R&D Monitoring (ECOOM, KU Leuven, Louvain, Belgium

² Faculty of Social Sciences, University of Antwerp, Antwerp, Belgium

³ Center for Science, Technology & Education Assessment (CSTEa), Wuhan University, Wuhan, China

⁴ School of Information Management, Wuhan University, Wuhan, China

⁵ Nordic Institute for Studies in Innovation, Research and Education, Oslo, Norway

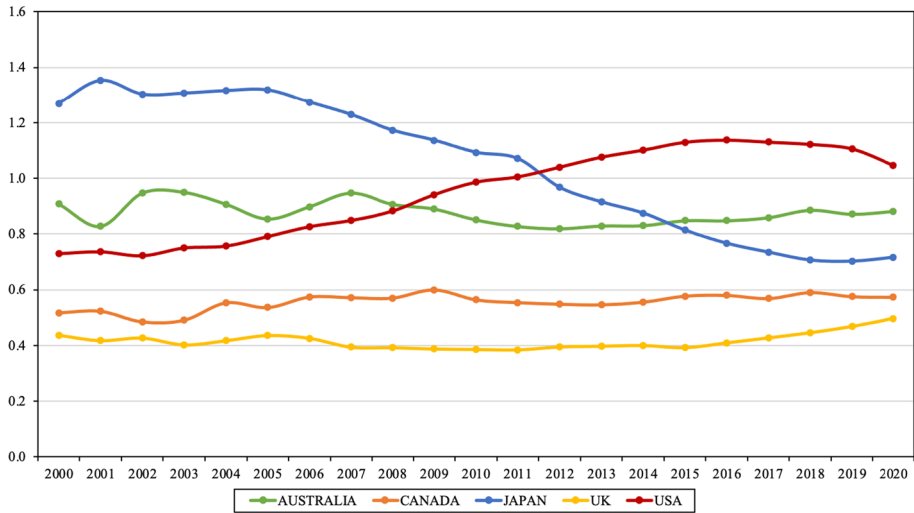


Fig. 1 The relative intensity of collaboration (RIC) over time from the perspective of China in relation to five major collaborating countries. China is X in RIC(X,Y)

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

A collaboration is counted each time author addresses in two different countries appear in the same publication. $C_{X,Y}$ denotes the number of collaborations between countries X and Y; C_X is the total number of collaborations country X has with all other countries, C_Y is the total number of collaborations country Y has with all other countries, and T is the total number of pairwise country collaborations in the set of publications under study.

As shown in Rousseau (2021), this formula may lead to inconsistent results. We presented a modification that leads to a more logical solution in Fuchs et al. (2021), see also (Fuchs & Rousseau, 2021) for a practical approach. A new collaboration indicator Relative Intensity for Collaboration (RIC) was introduced as:

$$RIC(X, Y) := \left(\frac{C_{XY}}{C_X}\right) / \left(\frac{C_Y - C_{XY}}{T - C_X}\right) = \frac{C_{XY} \cdot (T - C_X)}{C_X \cdot (C_Y - C_{XY})} \tag{2}$$

We refer to this formula as the relative intensity of collaboration of actor X to actor Y (in that order), actor X being the focus. We see that $RIC(X,Y) \geq 0$, but that there is no theoretical upper limit. This indicator is an asymmetric (the role of country X differs from the role of country Y) and a relative indicator. The main reason for preferring RIC over the indicator from 1992 (formula (1)) is that, ceteris paribus, when the collaboration between X and Y increases, RIC increases, which is not always the case for formula (1).

To demonstrate the properties of RIC with the use of real-world data, we used a matrix with collaboration frequencies among the 20 largest contributing countries¹ to journals indexed by the Web of Science in the years 2000–2020. Figure 1 provides an example where RIC is used to measure the relative intensity over time of China's collaboration with five major collaborating countries within the matrix of 20 countries.

The patterns seen in Fig. 1 are related to the growth of China within the global network. Ten years ago, China surpassed the United Kingdom as the USA's largest collaboration partner in science. The USA now has twice as many collaborations with China as with the UK. However, the relative intensity of collaboration between the USA and China has been declining since it peaked in 2016. China is seeking other partners, among them the UK. This recent development is only visible with an indicator of the relative intensity of collaboration. The decreased relative intensity of collaborations between China and the USA is possibly reflecting COVID-19 travel restrictions and the deteriorating relations between the two countries since 2018 (Tang et al., 2021; Zweig, 2021).

The policy interest in the patterns observed in Fig. 1 is obvious. The observations lead to other questions that may also be of policy interest: Is China developing a more balanced global collaboration profile as the intensity in relation to the USA decreases? And what happened to the balance of the global collaboration profile of the USA as China became the major collaboration partner?

This article introduces a new indicator to measure *the balance in collaboration profiles* (BIC). As a first step towards calculating the BIC, we prefer using the RIC as applied on a matrix of collaboration frequencies within a network, but similar probabilistic affinity indices used to study collaboration (Chinchilla-Rodríguez et al., 2021) may provide the basis as well. Then the weighted Lorenz curve is applied to measure the distances from a balanced relation. This procedure results in a weighted Gini index between 0 and 1. As the Gini index is widely used to measure *inequality* in the distribution of income across a population where 0 refers to perfect equality, we suggest turning it around and using the Gini evenness measure to express perfect balance (at 100%). This leads to the following definition.

Definition The balance of collaboration, denoted as BIC(L), of a country L with respect to a group of other countries, is equal to the evenness Gini index of the weighted Lorenz curve of collaboration, expressed as a percentage.

We explain further how such a weighted Lorenz curve of collaboration can be constructed and how the corresponding Gini evenness measure is calculated. The BIC can be used to compare actors in a network and how their collaboration profiles develop over time. We will demonstrate how the BIC is constructed with the example of the calculations of RIC in China's relations with 19 other countries with the same data as used in Fig. 1. The properties of the indicator will be discussed step by step. We end the analysis by presenting an example of BIC scores for China, the UK, and the USA based on the development within the same network during the period 2000–2020.

¹ See the list of countries in Table 3.

Methods

Introducing a weighted Lorenz curve of collaboration

Consider a set V of publications, representing e.g., a scientific field, originating from different countries. Moreover, we assume that each country in the set has at least one internationally collaborated publication. We focus on country L and count the number of times country L has publications, i.e., collaborated with each other country in the set. This is the number of collaboration links of each other country with country L . Although we prefer fractional counting [and there exist several options here, see e.g., (Rousseau & Zhang, 2021)], at this point it does not matter if one counts the number of publications using whole counting or using fractional counting. For simplicity, we formulate our theory in terms of whole counting. We recall that whole counting refers to participating, while fractional counting refers to the actual contribution to collaborative work.

If there are N other countries in the set V , this leads to a finite row $S = (s_1, \dots, s_N)$. If whole counting is used, s_j denotes the number of collaboration links between country L and country L_j , $j = 1, \dots, N$. Some components of S may be zero, but that is allowed in our approach.

Next, we count for each country, except L , the number of articles resulting from international collaboration (articles with more than one country in the byline) with countries different from L . Hence, for the other countries, we exclude collaborations with country L . Again, this can be done with whole numbers or fractionally (the fractional contribution of country L in each collaborated article). This leads to a finite row $R = (r_1, r_2, \dots, r_N)$. Indices must correspond of course: the index j in arrays S and R must refer to the same country. Because we assumed that each country has at least one internationally collaborated publication it does not occur that for a country the r - as well as the s -value is zero. It is possible though that one of the two values is zero. The s -value is zero if the country does not collaborate with country L , its r value is zero if this country only collaborates with country L .

Now the arrays S and R are normalized, leading to the arrays A and W with coordinates:

$$a_j = \frac{s_j}{\sum_{m=1}^N s_m} \quad \text{and} \quad w_j = \frac{r_j}{\sum_{m=1}^N r_m} \quad (3)$$

We note that $\sum_{m=1}^N s_m$ is equal to the total number of collaboration links between country L and all other countries. The sum $\sum_{m=1}^N r_m$ is equal to the total number of collaboration links between all countries, excluding country L . Yet, such links are now double counted as a collaboration between country L_i and L_j counts for country L_i and also for country L_j . Consequently $T = \sum_{m=1}^N s_m + \left(\sum_{m=1}^N r_m\right)/2$ is equal to the total number of collaboration links in set V .

Now we want to compare these two arrays. Does country L collaborate with other countries in proportion to their contribution in collaborations in field V , or does L has a few preferred countries?

We re-arrange the country values such that for the new arrangement

$$\frac{a_1}{w_1} \geq \frac{a_2}{w_2} \geq \dots \geq \frac{a_N}{w_N} \quad (4)$$

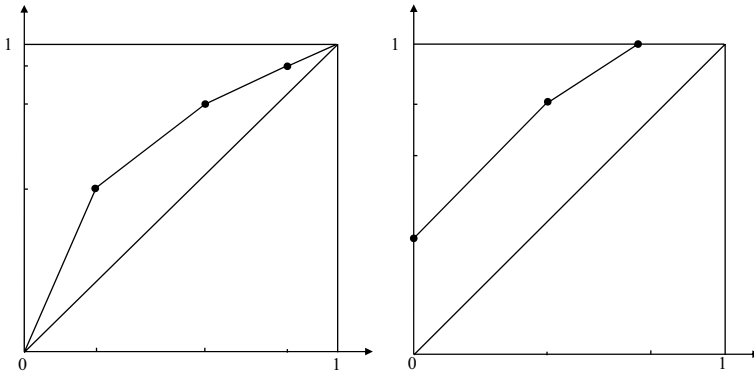


Fig. 2 Weighted Lorenz curves. **A** Standard discrete case. **B** Case with zero weights

where we have assumed that all components of W are different from zero. Now we can construct the corresponding weighted Lorenz curve. This is the broken line connecting the origin $(0,0)$ to the points with components

$$\left(\sum_{j=1}^i w_j, \sum_{j=1}^i a_j \right)_{i=1, \dots, N} \tag{5}$$

We easily see from its definition that if the normalized weight values are equal (to $1/N$) then we obtain the standard Lorenz curve (Lorenz, 1905; Rousseau et al., 2018, p. 88).

Note that the order of the ratios (a_j/w_j) already provides information about preferred countries for country L . If $a_j = w_j$ then country L collaborates with country j in the same proportion as country j has collaborated with other countries within this dataset. If $a_j > w_j$ then country L has a preference for collaborating with country L_j , while if $a_j < w_j$ the opposite is the case. Note that the ratios (a_j/w_j) are the slopes of the line segments of the weighted Lorenz curve. As these slopes decrease, see Eq. (4), the curve is concave. Only if all (a_j/w_j) -values are equal to 1, the weighted Lorenz curve coincides with the diagonal of the unit square. Two simple examples of the construction of a weighted Lorenz curve are shown in Fig. 2.

The usual theory of the weighted Lorenz curve does not include the case that some W -values are zero. If $w_1 = 0$, then $a_1/w_1 = \infty$ (recall that it is excluded that a_1 as well as w_1 are equal to zero). Yet, there is no real objection to beginning a weighted Lorenz curve with one or more vertical parts. Such a weighted Lorenz curve would be farther away from the diagonal than a similar one not beginning with a vertical part. This feature corresponds with our aim of comparing the A and W arrays.

A partial order among countries

We have explained the weighted Lorenz construction for a given country L . This construction can be repeated for all other countries. If we denote the weighted Lorenz curve of country L_j by $L_w(L_j)$, then we can introduce a partial order between countries, denoted as $-<$ as follows: $L_j -< L_m$ if $L_w(L_j) \leq L_w(L_m)$, with equality only if the two curves coincide. The relation $-<$ is only a partial order as curves may intersect, as they do for classical Lorenz curves.

The relation with RIC(X,Y)

Assuming that country X in Eq. (2), from Fuchs et al. (2021) is now the target country L and country Y is now L_j (the j th country), then $C_{X,Y}$ is now denoted as s_j ; C_X is now denoted as $\sum_{m=1}^N s_m$; C_Y being the total number of collaborations country Y has with all other countries except country X , is here denoted as r_j and $T - C_X$, the total number of pairwise country collaborations, not including those with country X , in the set of studied publications is now $(\sum_{m=1}^N r_m)/2$. Consequently, $w_j = \frac{r_j}{\sum_{m=1}^N r_m} = \frac{C_Y - C_{X,Y}}{2(T - C_X)}$ and $a_j = \frac{s_j}{\sum_{k=1}^N s_k} = \frac{C_{X,Y}}{C_X}$. Hence the slopes of the line segments of the weighted Lorenz curves, namely (a_j/w_j) are equal to twice the RIC(X,Y) values. When w_j is equal to zero, this is not a proper RIC value. It refers to the case that country L_j has an exclusive relation with country L .

Measures for global collaboration patterns

Now any measure respecting the partial order of the weighted Lorenz curve can be used as a collaboration measure. This means that if m denotes a measure and $L_j \prec L_n$ then $m(L_j) \leq m(L_n)$ with equality only if $L_j = L_n$. Probably the weighted Gini index is the easiest to understand. The interpretation of this index is the same as that of the (unweighted) Gini index, namely twice the area between the (weighted) Lorenz curve and the diagonal. This indicator can be calculated as:

$$G_w(L) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N |w_i a_j - w_j a_i| \quad (6)$$

or equivalently, when data are ranked as in Eq. (4), and $b_j = 1 - \sum_{m=1}^j a_m$ for $j = 1, \dots, N$, with $b_0 = 1$:

$$G_w(L) = 1 - \left(\sum_{i=1}^N b_{i-1} w_i + \sum_{j=1}^{N-1} b_j w_j \right) \quad (7)$$

Formula (6) can be found in Theil (1967, p. 121), while proof of formula (7) is given in the “Appendix”. This measure has been used e.g., in studies of the localization of industry under the name of ‘locational Gini coefficients’ (Krugman, 1991; Zitt et al., 1999).

If the first k weights are equal to zero, then this formula becomes (see “Appendix”):

$$G_w(L) = 1 - \sum_{j=k+1}^N (b_j + b_{j-1}) w_j \quad (8)$$

still with $b_j = 1 - \sum_{m=1}^j a_m$ for $j = 1, \dots, N$, and $b_0 = 1$.

Formulae (6), (7), and (8) are (classical) weighted Gini indices for inequality. As stated before, we need the weighted Gini evenness coefficient, which is equal to $1 - G_w(L)$. Hence, we have for countries X and Y :

$$BIC(X, Y) = \sum_{j=k+1}^N (b_j + b_{j-1}) w_j \quad (9)$$

The higher the weighted Lorenz curve the more country L 's collaboration pattern differs from the other countries' collaboration pattern.

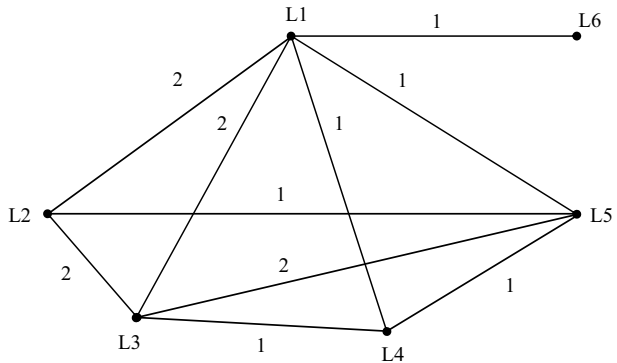
Table 1 Collaboration table (6 articles, 6 countries; we focus on country L_1)

	A1	A2	A3	A4	A5	A6
L_1	3	0	2	4	2	0
L_2	0	0	2	6	0	0
L_3	0	1	2	6	0	0
L_4	1	2	0	0	0	0
L_5	0	1	0	1	0	3
L_6	0	0	0	0	1	0

Table 2 Data for the case of collaborated articles using whole counting

	A1	A2	A3	A4	A5
L_1	1	0	1	1	1
L_2	0	0	1	1	0
L_3	0	1	1	1	0
L_4	1	1	0	0	0
L_5	0	1	0	1	0
L_6	0	0	0	0	1

Fig. 3 Weighted network corresponding with the data in Table 2



A simple fictitious case, illustrating the different steps in the whole counting case (the same idea can be applied to fractional counting) is provided in “[A fictitious illustrative example](#)” section. A real-world example follows in “[An example based on real data](#)” section.

A fictitious illustrative example

We consider the following collaboration Table 1 which refers to six articles and six countries (hence $N=5$), where we focus on country L_1 . The cells in the table refer to the number of authors of the country in the first column collaborating to the article in a given column.

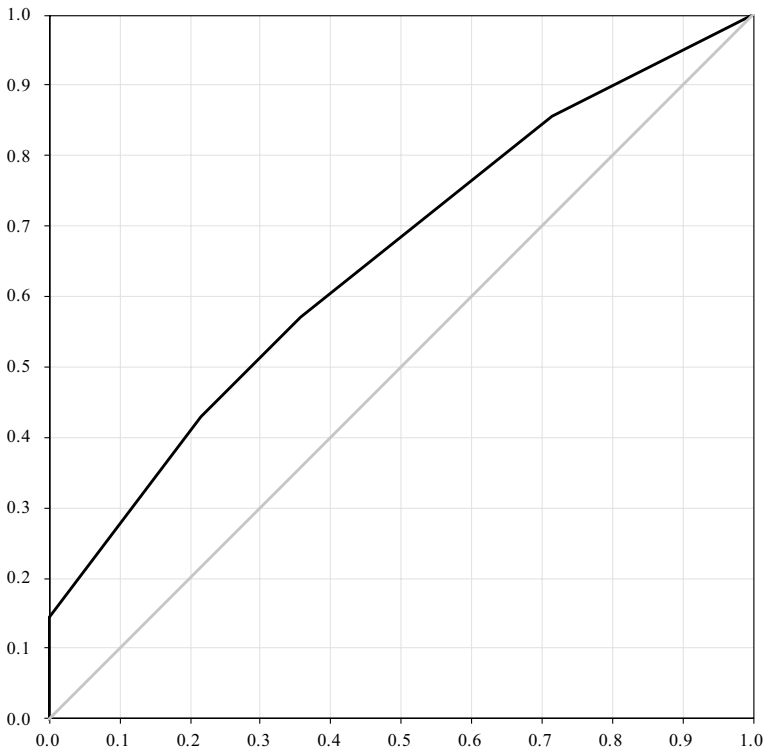


Fig. 4 Weighted Lorenz curve for country L with arrays $S=(2, 2, 1, 1, 1)$ and $R=(3,5,2,4,0)$

We will only consider internationally collaborated articles in this example (so A_6 is not considered). We will only illustrate the case of whole counting. Hence, we only use the data in Table 2.

Figure 3 shows the corresponding weighted network.

The array $S=(2, 2, 1, 1, 1)$ gives the number of articles in which country $L_j, j=2,3,4,5,6$, collaborates with country L_1 (with or without other countries). Normalizing leads to $A=(2/7, 2/7, 1/7, 1/7, 1/7)$.

The array R is here $(3,5,2,4,0)$, showing the number of collaboration links involving countries $L_j, j=2,3,4,5,6$. The corresponding normalized array is $W=(3/14, 5/14, 2/14, 4/14, 0/14)$.

Now the components of W and A must be re-arranged such that a_j/w_j is decreasing. This leads to L_6, L_2, L_4, L_3, L_5 with corresponding (a_j/w_j) values: $+\infty > 4/3 > 1 > 4/5 > 2/4$. Recall that for the finite results, these slopes are equal to twice the $\text{RIC}(L_1, L_j)$ values.

Now we can draw the weighted Lorenz curve. It connects the points with components $(0,0) \rightarrow (0, 1/7) \rightarrow (3/14, 3/7) \rightarrow (5/14, 4/7) \rightarrow (10/14, 6/7) \rightarrow (1,1)$. It is always concave. The result for this example is shown in Fig. 2.

The value of the weighted Gini evenness index is $34/49$. This means that the BIC is equal to 69.4%.

We note that using $S=(200,200,100,100,100)$ and $R=(300,500,200, 400,0)$ would lead to the same weighted Lorenz curve and hence also to the same BIC value (Fig. 4).

Table 3 A basis for the Lorenz curve

Country/region Y	$C_{X,Y}$ - array = Array S	$(C_Y - C_{X,Y})$ array = Array R	Array A	Array W	RIC = $A/(2W)$
USA	47,911	227,349	0.375	0.167	1.122
Taiwan	3701	19,028	0.029	0.014	1.015
Australia	11,334	65,111	0.089	0.05	0.886
Japan	6704	46,590	0.052	0.037	0.706
South Korea	4014	31,554	0.031	0.026	0.613
Canada	8069	65,544	0.063	0.054	0.59
United Kingdom	12,505	130,437	0.098	0.11	0.446
India	2340	30,020	0.018	0.026	0.355
Russia	2119	29,313	0.017	0.025	0.328
Germany	7473	113,442	0.058	0.099	0.296
Sweden	2577	39,933	0.02	0.035	0.29
Turkey	864	14,977	0.007	0.013	0.257
France	4817	83,808	0.038	0.074	0.256
Netherlands	2960	56,104	0.023	0.049	0.234
Poland	1275	27,365	0.01	0.024	0.205
Italy	3071	73,365	0.024	0.065	0.184
Brazil	1399	33,519	0.011	0.03	0.183
Spain	2547	61,150	0.02	0.055	0.183
Switzerland	2115	53,468	0.017	0.048	0.173
Sum	127,795	1,202,077	1	1	

Data ranked according to descending RIC values in 2018 for China in relation to 19 other countries

An example based on real data

To provide examples based on real-world data, we used *Web of Science* as represented in Clarivate’s *InCites* in 2021 to register co-occurrences in the addresses of scientific articles (including reviews) of the 20 countries listed in Table 4 below. A co-occurrence is counted only once even if a country is listed more than once in an article. The 20 countries were selected by the size of their scientific output and their size within the global network of international collaboration. Together, the 20 countries contributed to 87.6% of the world’s scientific output in publications in 2020. Our measurement will not change much by adding more countries. However, the aim of collaboration analysis is often to compare countries of the same size or within the same region. We suggest adding smaller countries to a group of large countries rather than replacing large countries with smaller countries if the intention is to provide an unbiased picture of how smaller countries collaborate within the global collaboration network.

To demonstrate how the BIC is constructed using a real-world example, we consider the collaboration pattern of China in the year 2018 with the other 19 countries/regions. To create the Lorenz curve, the other countries need to be ranked by descending RIC as shown in Table 3. The basis for the Lorenz curve is the two columns Array A and Array W which are also the basis for calculating the RIC. To take the example of China in relation to the USA, the RIC value of 1.122 is calculated as the observed relative intensity (0.375) divided

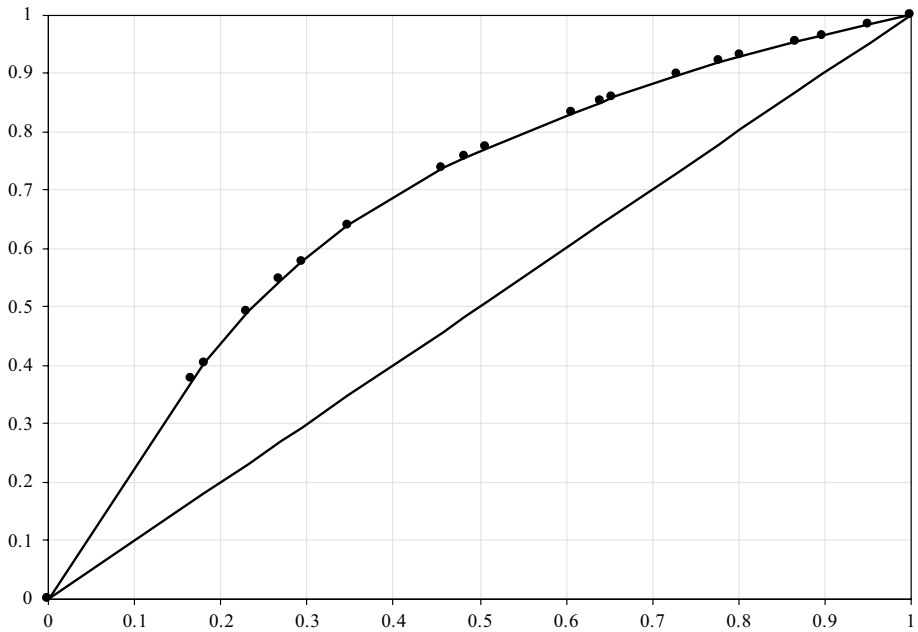


Fig. 5 Weighted Lorenz curve for China in 2018

by the expected relative intensity (0.334). The first of these values is used in Array *A* while half of the latter is used in Array *W*. After the columns are made cumulative, Array *A* provides the y-axis and Array *W* the x-axis for the Lorenz curve. Array *W* provides the weighing of the *A* values. The result is shown in Fig. 5. The value of the weighted Gini evenness index is 0.637. This means that the BIC is equal to 63.7%.

Two examples of the use of the BIC

Before concluding, we will use the same dataset to provide two examples of how the BIC can be used in policy-relevant analysis. Figure 6 shows the development of the BIC in the period 2000–2020 for the same six countries that are represented in Fig. 1 above. The UK stands out with a high degree of balance in international scientific collaboration while China stands out with a low degree. The general trend is towards increased balance in international scientific relations, but this is the case for China and the USA only in recent years. When comparing with Fig. 1, it seems that the increasingly intense collaboration between the two countries until 2016 came at a loss of balance and that the reduced bilateral intensity in the most recent years has provided for increased balance in the two countries' relations to other countries.

The BIC is probably most relevant to use in combination with RIC or similar measures of collaboration intensity. An example is given in Table 4. Here, we have ranked the 20 countries in descending order according to the BIC values in 2020. In the other columns, we identify the countries or groups of countries that exhibit the highest and lowest RIC values from the perspective of each country in the same year. The relative

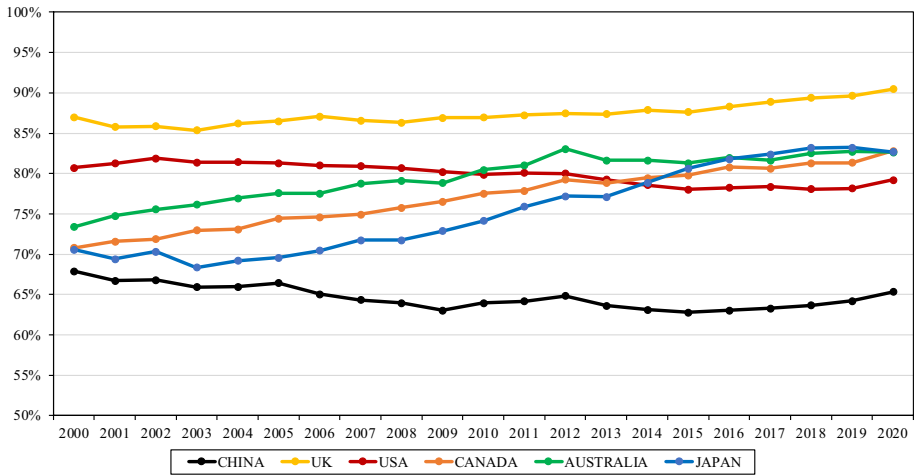


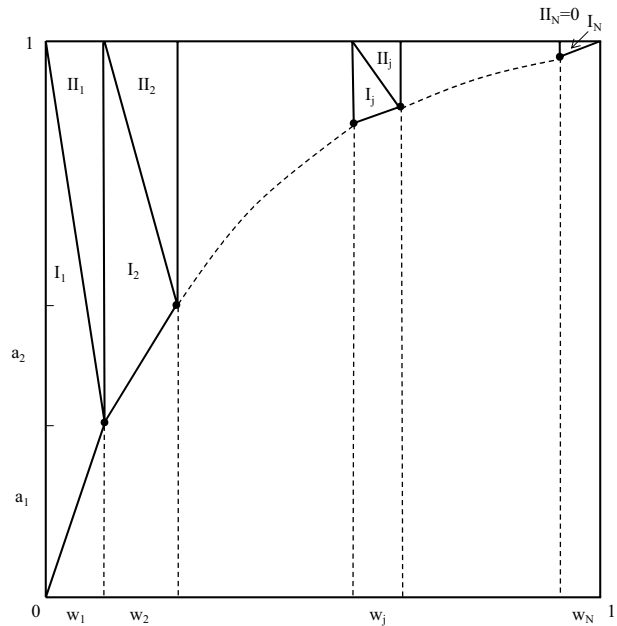
Fig. 6 Example of BIC values: The balance in the collaboration patterns of China and its five major collaborating countries in the years 2000–2020

Table 4 The 20 countries ranked according to BIC in 2022 with additional information about the bilateral relations with the highest and lowest collaboration intensity according to RIC

Country/region Y	BIC (%)	High RIC	Low RIC
United Kingdom	90.5	Europe and Australia	Taiwan, South Korea, Japan, Russia
Sweden	86.3	Europe	Asia
Turkey	84.2	Taiwan, India, Russia, Poland	Canada, China, Australia, Japan
Germany	83.3	Europe	Asia, Australia, Canada
France	82.9	Europe	Asia, Australia
Canada	82.8	USA, China, Brazil, Australia	Russia, Turkey, Taiwan, Poland
Japan	82.7	Asia, USA	Europe
Australia	82.6	China, Commonwealth, USA	Europe, Brazil
Russia	82.5	Poland, Turkey, Germany	Commonwealth, USA, China
Brazil	81.6	Spain, USA, France, India, Canada	Other Europe, Asia
India	81.5	South Korea, Taiwan, Turkey, Russia	
Netherlands	81.2	Europe	Asia, Russia, Brazil
Poland	79.2	Russia, Turkey, Italy, Spain	China, Australia, Canada, USA
USA	79.2	China, South Korea, Canada	Poland, Russia, Europe
Switzerland	78.1	Europe	Asia, Australia
Italy	77.9	Europe	Asia, Australia
Spain	77.6	Europe, Brazil	Asia
South Korea	70.4	Asia, USA	Europe
Taiwan	67.4	Asia, Turkey	Europe
China	65.3	USA, Asia, Australia, Canada, UK	Europe, Brazil

intensities could also have been visualized in a two-dimensional map, but here, the BIC adds information about the balance.

Fig. 7 Calculation of the weighted Gini index



Conclusion

The BIC indicator is a new indicator, specifically constructed to measure balance in collaboration. The indicator is based on the Gini index for a weighted Lorenz curve. The relation between this weighted Lorenz curve and the previously introduced indicator of relative intensity of collaboration (RIC) is explained. As the weighted Lorenz curve and the associated Gini index are not often used in bibliometric work, their construction and calculation are clearly explained.

Our examples using a dataset with collaboration frequencies among 20 large countries demonstrate that the new indicator of balance in collaboration patterns can be a useful extension of indicators measuring the relative intensity of collaboration in bilateral relations. For any country, the balance can be studied over time and compared to other countries.

Appendix: Proof of formula (11) and its equality with formula (10)

Weighted Gini evenness index

We calculate the area above the weighted Lorenz curve. Then the weighted Gini evenness index is twice the calculated area. For the BIC indicator, we use this value expressed as a percentage.

The area above the weighted Lorenz curve is the sum of two types of triangles, those indicated with I and those indicated with II, see Fig. 7. Recall that $b_j = 1 - \sum_{k=1}^j a_k$ with $j = 1, \dots, N$, and $b_0 = 1$.

Twice the area of I_1 is: $w_1 \cdot 1 = w_1 \cdot b_0$.
 Twice the area of Π_1 is: $w_1 \cdot (1 - a_1) = w_1 \cdot b_1$.
 Twice the area of I_2 is: $w_2 \cdot (1 - a_1) = w_2 \cdot b_1$.
 Twice the area of Π_2 is: $w_2 \cdot (1 - (a_1 + a_2)) = w_2 \cdot b_2$.
 In general: twice the area of I_j is: $w_j \cdot b_{j-1}$.
 Twice the area of Π_j is: $w_j \cdot b_j$.
 Twice the area of I_N is: $w_N \cdot b_{N-1}$.
 Twice the area of Π_N is zero.

Hence the area above the weighted Lorenz curve is

$$\frac{1}{2} \left(\sum_{j=1}^N w_j b_{j-1} + \sum_{j=1}^{N-1} w_j b_j \right)$$

Finally, the Gini evenness coefficient is:

$$\left(\sum_{j=1}^N w_j b_{j-1} + \sum_{j=1}^{N-1} w_j b_j \right)$$

If there are k zero weights, then two times I_1 becomes:

$$\left(1 - \sum_{m=1}^k a_m \right) w_{k+1} = b_k w_{k+1}$$

and the formula becomes:

$$\left(\sum_{j=k+1}^N w_j b_{j-1} + \sum_{j=k+1}^{N-1} w_j b_j \right)$$

still with $b_j = 1 - \sum_{m=1}^j a_m$ for $j = 1, \dots, N$, and $b_0 = 1$.

Proposition. The following two formulae (10) and (11) yield the same result.

$$G_w(L) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N |w_i a_j - w_j a_i| \tag{10}$$

$$\text{and } G_w(L) = 1 - \left(\sum_{i=1}^N b_{i-1} w_i + \sum_{j=1}^{N-1} b_j w_j \right) \tag{11}$$

with $b_j = 1 - \sum_{m=1}^j a_m$ with $j = 1, \dots, N$, and $b_0 = 1$.

Proof We first note that for $j > 0$: $\sum_{m=1}^j a_m = 1 - b_j$ and $\sum_{m=j+1}^N a_m = b_j$. Now,

$$\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N |w_i a_j - w_j a_i| = \sum_{\substack{i,j \\ i < j}} (a_i w_j - a_j w_i) \left(\text{as } \frac{a_i}{w_i} \geq \frac{a_j}{w_j} \right).$$

$$\begin{aligned}
&= (a_1w_2 - a_2w_1 + a_1w_3 - a_3w_1 + a_1w_4 - a_4w_1 + \dots + a_1w_N - a_Nw_1) \\
&\quad + (a_2w_3 - a_3w_2 + a_2w_4 - a_4w_2 + a_2w_5 - a_5w_2 + \dots + a_2w_N - a_Nw_2) \\
&\quad + (a_3w_4 - a_4w_3 + a_3w_5 - a_5w_3 + \dots + a_3w_N - a_Nw_3) \\
&\quad + \dots + a_{N-1}w_N - a_Nw_{N-1} \\
&= w_1(-a_2 - a_3 - \dots - a_N) + w_2(a_1 - a_3 - a_4 - \dots - a_N) + w_3(a_1 + a_2 - a_4 - a_5 - \dots - a_N) \\
&\quad + w_N(a_1 + a_2 + \dots + a_{N-1}) \\
&= w_1(1 - b_0 - b_1) + w_2(1 - b_1 - b_2) + w_3(1 - b_2 - b_3) + \dots + w_N(1 - b_{N-1})
\end{aligned}$$

$$= 1 - \sum_{j=1}^N (b_{j-1} + b_j)w_j \text{ (recall that } b_0=1 \text{ and } b_N=0)$$

$$\text{Now, } G_w(L) = 1 - \left(\sum_{i=1}^N b_{i-1}w_i + \sum_{j=1}^{N-1} b_jw_j \right) = 1 - \sum_{j=1}^N (b_j + b_{j-1})w_j$$

$$\text{If the first } k \text{ weights are zero, then we have: } G_w(L) = 1 - \sum_{j=k+1}^N (b_j + b_{j-1})w_j.$$

Acknowledgements The authors thank Ying Huang for providing the China-USA dataset. Lin Zhang would like to acknowledge support from the National Natural Science Foundation of China (Grant Nos. 71974150 and 72074029), and the National Laboratory Center for Library and Information Science in Wuhan University. Gunnar Sivertsen acknowledges support from the Research Council of Norway, Grant 256223, FORINNPOL.

Declarations

Competing interest Ronald Rousseau is a member of the Distinguished Reviewer Board of Scientometrics, and Lin Zhang is the co-editor-in-chief of Scientometrics.

References

- Chinchilla-Rodríguez, Z., Bu, Y., Robinson-García, N., & Sugimoto, C. (2021). An empirical review of the different variants of the probabilistic affinity index as applied to scientific collaboration. *Scientometrics*, *126*, 1775–1795.
- Fuchs, J. E., & Rousseau, R. (2021). How to calculate the relative intensity of collaboration (RIC) for countries from Web of Science data. *ISSI Newsletter*, *17*(3), 24–29.
- Fuchs, J. E., Sivertsen, G., & Rousseau, R. (2021). Measuring the relative intensity of collaboration within a network. *Scientometrics*, *126*(10), 8673–8682.
- Krugman, P. (1991). *Geography and trade*. University Press.
- Lorenz, M. O. (1905). Methods of measuring concentration of wealth. *Journal of the American Statistical Association*, *9*, 209–219.
- Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology, & Human Values*, *17*(1), 101–126.
- Rousseau, R. (2021). Going back in time: Understanding patterns of international scientific collaboration. *Journal of Scientometric Research*, *10*(1), 126–129.
- Rousseau, R., & Zhang, L. (2021). Bilateral co-authorship indicators based on fractional counting. *Journal of Data and Information Science*, *6*(1), 1–12.
- Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming metric-wise. A bibliometric guide for researchers*. Kidlington (Chandos-Elsevier).
- Tang, L., Cao, C., Wang, Z., & Zhou, Z. (2021). Decoupling in science and education: A collateral damage beyond deteriorating US–China relations. *Science and Public Policy*, *48*(5), 630–634.
- Theil, H. (1967). *Economics and information theory*. North-Holland.
- Zitt, M., Barré, R., Sigogneau, A., & Laville, F. (1999). Territorial concentration and evolution of science and technology activities in the European Union: A descriptive analysis. *Research Policy*, *28*, 545–562.
- Zweig, D. A. (2021). Is sino-american scientific collaboration a thing of the past? *International Higher Education*, *108*, 5–7.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.