Results presented in:
L. Egghe and R. Rousseau, Introduction to Informetrics
Quantitative Methods in Library, Documentation,
and Information Science (Amsterdam: Elsevier, 1990)
pp. 274 - 280

COMPARING TWO WEIGHING METHODS IN CITATION ANALYSIS

Eda Kranakis

Wetenschappsdynamica, Universiteit van Amsterdam
Amsterdam, The Netherlands

Evangelos Kranakis

Centrum voor Wiskunde en Informatica
Kruislaan 413, Amsterdam, The Netherlands
(eva@cwi.nl)

ABSTRACT

We study the behavior of two weighing methods: Unit Counting and Fractional Counting, used in citation analysis. This indicates that results obtained using these two methods can contradict each other.

1. Two Weighing Methods

Let us first establish the following notation.

C = is a set of countries,

n = is the number of articles,

a_i = is the number of co-authors with contributions to the i th article, and

$a_i(c)$ = is the number of co-authors from country c with contributions to the i th article.

It is obvious that for all $i = 1,2,...,n$,

$$a_i = \sum_{c \in C} a_i(c).$$

We interested in comparing the following two ways of counting the total contribution of a country c in the given set of n papers.

1. (**Unit Counting**) Every co-author's contribution in each article weighs exactly 1 point.

In this case, the total weight (i.e., number of points) assigned is

$$W = \sum_{i=1}^{n} a_i .$$

The total weight (i.e., number of points) assigned to country c is

$$W(c) = \sum_{i=1}^{n} a_i(c).$$

And the contribution of country c in the overall assignment of points is

$$Q(c) = \frac{W(c)}{W} = \frac{\sum_{i=1}^{n} a_i(c)}{\sum_{i=1}^{n} a_i}.$$

2. (**Fractional Counting**) Every co-author's contribution in the i th article weighs exactly $1/a_i$ points. In which case, the weight of country c in the i th article must be $a_i(c)/a_i$.

In this case, the total weight (i.e., number of points) assigned is

$$W' = \sum_{1 \le i \le n} \sum_{c \in C} \frac{a_i(c)}{a_i} = n.$$

The total weight (i.e., number of points) assigned to country c is

$$W'(c) = \sum_{1 \le i \le n} \frac{a_i(c)}{a_i}.$$

And the contribution of country c in the overall assignment of points is

$$Q'(c) = \frac{W'(c)}{W'} = \frac{\sum_{i=1}^{n} a_i(c)/a_i}{n}.$$

2. Comparison of the Two Weighing Methods

We are interested in comparing the quantities $Q(c)$, $Q'(c)$. Put

$\Delta(c) = |Q(c) - Q'(c)|$,

$m = \min(a_1, \ldots, a_n)$,

$M = \max(a_1, \ldots, a_n)$.

It is then easy to show that the following theorem holds.

Theorem. For any country c,

$$\Delta(c) = |Q(c) - Q'(c)| \le \left(\frac{1}{m} - \frac{1}{M} \right) \cdot \frac{\sum_{i=1}^{n} a_i(c)}{n},$$

where m (respectively, M) is the minimum (respectively, maximum) number of co-authors in an article from the given list of n articles.

Proof.

To see this, put

$$Q_i(c) = \frac{a_i(c)}{a_i}$$

and notice that

$$Q'(c) = \frac{1}{n} \cdot [Q_1(c) + \cdots + Q_n(c)].$$

However, for all $= 1, \ldots, n$,

$$\frac{a_i(c)}{M} \le Q_i(c) \le \frac{a_i(c)}{m}.$$

Hence, from the definitions of $Q(c)$, $Q'(c)$ we obtain that

$$\frac{1}{n\cdot M}\cdot\sum_{i=1}^{n}a_i(c)\leq Q(c),\ Q'(c)\leq\frac{1}{n\cdot m}\cdot\sum_{i=1}^{n}a_i(c),$$

from which the theorem follows easily. •

2.1. Examples.

1. If all the a_i's are equal then $m = M$ and hence the theorem implies that $Q(c) = Q'(c)$.
2. If for all c there exists a constant $\lambda(c)$ such that for all $i = 1,\ldots,n$, $a_i(c)/a_i = \lambda(c)$ then $Q(c) = Q'(c)$. This shows that the upper bound given in the theorem is not optimal.
3. The quantities $Q(c), Q'(c)$ can diverge. For example, fix c and consider the following assignment of values: $a_i(c) = 1$, $a_i = i$, where $i = 1,\ldots,n$. Then it is clear that

$$W = 1 + 2 + \cdots + n = n(n+1)/2,\quad W' = n,$$

$$W(c) = n,\quad W' = \frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}\approx\log n,$$

Hence, $Q(c) = \frac{2}{n+1}$, and $Q'(c)\approx\frac{\log n}{n}$. Consequently,

$$\Delta(c)\approx\left|\frac{2}{n+1}-\frac{\log n}{n}\right|,$$

and

$$\frac{Q(c)}{Q'(c)}\approx\frac{2}{\log n}.$$

2.2. A Two Country Example

In this example we make the following assumptions (see table 1):

A1. $C = \{c, c'\}$, i.e. there are two countries,

A2. $1\leq a_i(c)\leq a_i\leq 2$, i.e. country c contributes on every article and every article has either one or two authors (but no more),

A3. there exist integers k, l such that

- $a_1 = \cdots = a_k = 1$, i.e. the first k articles have exactly one author,
- $a_{k+1} = \cdots = a_n = 2$ (i.e. the next $n-k$ articles have exactly two authors) and among these articles the first l have exactly one author from country c (i.e. $a_{k+1}(c) = \cdots = a_{k+l}(c) = 1$) and the remaining $n-k-l$ have exactly two authors from c (i.e. $a_{k+l+1}(c) = \cdots = a_n(c) = 2$).

Now for the model described above we have the following identities:

$$Q(c) = \frac{a_1(c)+\cdots+a_n(c)}{a_1+\cdots+a_n} = \frac{k+l+2(n-k-l)}{k+2(n-k)} = 1-\frac{l}{2n-k},$$

$$Q'(c) = \frac{1}{n}\cdot\left[\frac{a_1(c)}{a_1}+\cdots+\frac{a_n(c)}{a_n}\right] = \frac{1}{n}\cdot\left[k+\frac{l}{2}+n-k-l\right] = 1-\frac{l}{2n}.$$

It follows that

a_1	$\cdots$	a_k	a_{k+1}	$\cdots$	a_{k+l}	a_{k+l+1}	$\cdots$	a_n
1	$\cdots$	1	2	$\cdots$	2	2	$\cdots$	2
$a_1(c)$	$\cdots$	$a_k(c)$	$a_{k+1}(c)$	$\cdots$	$a_{k+l}(c)$	$a_{k+l-1}(c)$	$\cdots$	$a_n(c)$
1	$\cdots$	1	1	$\cdots$	1	2	$\cdots$	2
$a_1(c')$	$\cdots$	$a_k(c')$	$a_{k+1}(c')$	$\cdots$	$a_{k+l}(c')$	$a_{k+l+1}(c')$	$\cdots$	$a_n(c')$
0	$\cdots$	0	1	$\cdots$	1	0	$\cdots$	0

Table 1: The Two Country Model

$$\Delta(c) = \frac{lk}{2n(2n-k)}.$$

Now assuming that $l + k = n$, the quantity $\Delta(c)$ obtains its maximum value

$$\max_{l+k=n} \Delta(c) = \frac{3\sqrt{2}-4}{2\sqrt{2}} \approx 0.09,$$

when

$$k = 2n - \sqrt{2}n, \; l = \sqrt{2}n - n.$$

In other words, for the above mentioned values of k, l there is a 9% difference between $Q(c), Q'(c)$.

The following qu?stion arises:

Question:

Are there any conditions under which the % contribution of a country will appear to decline because their rate of international co-authorship is going up?

Example.

This is rather easy to see in the previous example. Indeed,

$$Q(l,k,n,c) = 1 - \frac{l}{2n-k}, \; Q'(l,k,n,c) = 1 - \frac{l}{2n},$$

where l is the number of co-authored articles, and k is the number of articles with a single author (from country c). It is then clear that

$$Q'(n/4, k, n, c) = 1 - \frac{1}{8} = \frac{7}{8},$$

while

$$Q'(n/2, k, 3n/2, c) = 1 - \frac{1}{6} = \frac{5}{6}.$$

Hence, although the number of articles co-authored by country c is going up its percentage contribution (according to the fractional counting method) is going down! In the first case 25% of the articles are co-authored (i.e. $n/4$ out of the total n), while in the second 1/3 of the articles are co-authored (i.e. $n/2$ out of the total $3n/2$). A similar example can be given for the unit counting method.

2.3. The Two Country Example Revisited

Here we give the more general version of the previous example. We make the following assumptions:

A1. C is an arbitrary set of countries,

A2. for any country c, $0 \le a_i(c) \le a_i \le 2$, i.e. every article has either one or two authors (but no more),

A3. there exist integers k, l, s, t such that

- $a_1 = \cdots = a_k = 1$,

  $a_1(c) = \cdots = a_l(c) = 0$,

  $a_{l+1}(c) = \cdots = a_k(c) = 1$,

- $a_{k+1} = \cdots = a_n = 2$,

  $a_{k+1}(c) = \cdots = a_{k+s}(c) = 0$,

  $a_{k+s+1}(c) = \cdots = a_{k+s+t}(c) = 1$,

  $a_{k+s+t+1}(c) = \cdots = a_n(c) = 2$.

Clearly, $t = t(c)$ is the number of articles co-authored by authors from country c, As before, it is easy to obtain the following formulas:

$$Q(c) = \frac{a_1(c) + \cdots + a_n(c)}{a_1 + \cdots + a_n} = 1 - \frac{l + t + 2s}{2n - k},$$

$$Q'(c) = \frac{1}{n} \cdot \left[\frac{a_1(c)}{a_1} + \cdots + \frac{a_n(c)}{a_n} \right] = 1 - \frac{2l + t + 2s}{2n}.$$

3. The General Case

Here we give the more general case. We make the following assumptions:

A1. C is an arbitrary set of countries,

A2. for any country c, $0 \le a_i(c) \le a_i \le s$, i.e. every article has at most s authors.

A3. there exist integers $n_1, n_2, \ldots, n_{s-1}$ such that $n_1 + n_2 + \ldots + n_{s-1} \le n$, and

$a_1 = \cdots = a_n = 1$,

$a_{n_1+1} = \cdots = a_{n_2} = 2$,

...

$a_{n_{s-1}+1} = \cdots = a_n = s$.

It follows that

$$a_1 + \cdots + a_n = n_1 + 2(n_2 - n_1) + 3(n_3 - n_2) + \cdots + s(n - n_{s-1}) = sn - n_1 - n_2 - \cdots - n_{s-1}.$$

Next define the following quantities:

$cl(c) =$ number of collaborations of country c,

$cl(i, c) =$ number of collaborations of country c in articles with a total of i many authors,

$cm(c) =$ number of articles written only by authors from country c, each article counted by the number of authors it has,

$cm'(c) =$ number of articles written only by authors from country c, each article counted only once.

It is then clear that

$$cl(1,c) = 0,\ cl(c) = cl(2,c) + \cdots + cl(s,c).$$

$$Q(c) = \frac{cl(2,c) + \cdots + cl(s,c) + cm(c)}{sn - n_1 - \cdots - n_{s-1}}$$

$$Q'(c) = \frac{1}{n} \cdot \left[\frac{cl(2,c)}{2} + \cdots + \frac{cl(s,c)}{s} + cm'(c) \right]$$