Scientometrics, Vol. 3. No. 5 (1981) 363–377

THE DISTRIBUTION OF CITATIONS FROM NATION TO NATION ON A FIELD BY FIELD BASIS – A COMPUTER CALCULATION OF THE PARAMETERS

CLARE E. BURKE*, D. de S. PRICE**

**Worcester Polytechnic Institute, Worcester, Massachusetts, MA 01609 (USA)*
***Yale University, History of Science, Box 2036, Yale Station, New Haven, Conn. 06520, (USA)*

(Received July 3, 1980 in revised form March 6, 1981)

Following the methodology established by *Price,* this paper analyzes the empirical evidence of citation matrices. Using the data cleaned and tabulated by Computer Horizons, Inc. from the Science Citation Index data banks, it is shown that the non-diagonal elements of the square citation matrices can be accounted for very satisfactorily by assigning each nation a characteristic output and input coefficient in each field measured; the ratio of these coefficients provides a measure of quality. Deviations from this simple model give measures of particular linkage strengths between nations showing some evidence of preferences and avoidances that exist for reason of language, social structure, etc. It is also shown that the diagonal data can be accounted for by the measurable phenomenon that each nation seems to publish partly for the international knowledge system and partly for its own domestic purposes. Thus, three parameters and a cluster map can parsimoniously describe the citation data within the limits of random error.

Undoubtedly the largest scientometric data-base in existence is that provided as an accidental by-product of the production of the *Science Citation Index.*[1] There has recently become available a carefully cleaned and tabulated nation by nation and field by field analysis of such citations prepared by Computer Horizons Inc.[2] in connection with the contractual provision of data for the *Science Indicators* volumes.[3] Now that this material is available it satisfies a long felt demand for a large sample adequately disaggregated by nation and by field. The analysis has been shown by *Price* to be fairly straight-forward. Each element of the transactional matrix has a value close to that predicted as the product of two parameters, an output coefficient and an input coefficient for each of the nations. The exception only is the diagonal of the matrix, since the self citations of a nation to itself are evidently on a very different basis from those between dissimilar nations. Following *Price's* method we temporarily discard these diagonal values and then reinsert them later to separate that part of the citation flow which is international from that part which is "domestic".

Table 1
Original citation matrix for Physics, 1973

		Cited country							
		US	UK	WG	FR	SS	JA	CA	OT
Citing country	US	49 570	4 094	2 793	2 296	1 230	1 614	1 909	7 001
	UK	5 608	4 875	630	513	180	360	408	1 678
	WG	5 072	860	3 814	537	288	471	304	1 907
	FR	3 591	700	490	2 953	233	334	316	1 294
	SS	3 840	560	519	449	6 950	556	220	1 501
	JA	4 233	647	489	388	283	3 835	257	1 185
	CA	2 760	560	303	295	66	199	1 996	874
	OT	11 922	2 502	1 927	1 463	804	956	975	12 038

US – United States; UK – United Kingdom; WG – West Germany; FR – France; SS – Soviet Union; JA – Japan; CA – Canada; OT – Other countries

We illustrate the procedure by taking a single sample, the field of Physics (1973 data), and following this from the original data to the final reduction to parameters, the whole operation having been computer programmed in BASIC and run on the Worcester Area Community Center at Worcester Polytechnic Institute.[5]

The original Computer Horizons data drawn after much editing from the Science Citation Index Tapes is shown in Table 1. It is in fact one out of nine separate major fields into which all scientific literature had been subdivided, and data of this sort was obtained both for 1973 and 1974. For a fuller account of definitions, corrections and limitations of this data see M. *Carpenter.*[6] Each table shows a set of eight nations (actually seven major countries and a category of "Other Nations" into which the rest of the world was grouped) arranged both horizontally as "cited country" and vertically as "citing country". The first step in computation was a program for replacing the diagonal terms in this matrix by those which would be there if each nation cited itself in the same proportion as it cited all other nations. This was achieved by performing simple mathematical operations in a different sequence than described by *Price.* The objective was to avoid large numbers that the computer could not handle.

Diagonal reductions are used to adjust the numbers located on the diagonal. This will reduce the self citations so that they will be in proportion to the rest of the matrix. The reduction is accomplished in five steps.

In the first step any elements equal to zero are replaced with unity. All diagonal elements are also replaced with unity. This matrix is named $\overline{A}$.

New matrix generated by Step 1

$$
\bar{A} = \begin{bmatrix}
1 & 4094 & 2793 & 2296 & 1230 & 1614 & 1909 & 7001 \\
5608 & 1 & 630 & 513 & 180 & 360 & 408 & 1678 \\
5072 & 860 & 1 & 537 & 288 & 471 & 304 & 1907 \\
3591 & 700 & 490 & 1 & 233 & 334 & 316 & 1294 \\
3840 & 560 & 519 & 449 & 1 & 556 & 220 & 1501 \\
4223 & 649 & 489 & 388 & 283 & 1 & 257 & 1185 \\
2760 & 560 & 303 & 295 & 66 & 199 & 1 & 874 \\
11922 & 2502 & 1927 & 1463 & 804 & 956 & 975 & 1
\end{bmatrix}
$$

In the second step the computer program generates a new matrix, $\bar{B}$, by raising all elements of $\bar{A}$ to the one-sixth (0.16667) power ($B_{ij} = A_{ij}^{0.16667}$). This matrix is shown below.

New matrix generated by Step 2

$$
\bar{B} = \begin{bmatrix}
1 & 4.0 & 3.8 & 3.6 & 3.3 & 3.4 & 3.5 & 4.4 \\
4.2 & 1 & 2.9 & 2.8 & 2.4 & 2.7 & 2.7 & 3.4 \\
4.1 & 3.1 & 1 & 2.9 & 2.6 & 2.8 & 2.6 & 3.5 \\
3.9 & 3.0 & 2.8 & 1 & 2.5 & 2.6 & 2.6 & 3.3 \\
4.0 & 2.9 & 2.8 & 2.8 & 1 & 2.9 & 2.5 & 3.4 \\
4.0 & 2.9 & 2.8 & 2.7 & 2.6 & 1 & 2.5 & 3.3 \\
3.7 & 2.9 & 2.6 & 2.6 & 2.0 & 2.4 & 1 & 3.1 \\
4.8 & 3.7 & 3.5 & 3.4 & 3.0 & 3.1 & 3.1 & 1
\end{bmatrix}
$$

In step three matrices $\bar{R}$ and $\bar{C}$ are generated from B. Each element in $\bar{R}$ corresponds to the product of all the elements in one row of $\bar{B}$, while each element of $\bar{C}$ corresponds to the product of all the elements in one column of $\bar{B}$.

Step 3

$$
\bar{R} = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \\ R_6 \\ R_7 \\ R_8 \end{bmatrix}
$$

$$
\begin{aligned}
\text{Where } R_i &= B_{i1} \times B_{i2} \times B_{i3} \ldots B_{i8} \\
\text{Therefore } R_1 &= 1 \times 4.0 \times 3.8 \ldots 4.4 \\
R_2 &= 4.2 \times 1 \times 2.9 \ldots 3.4
\end{aligned}
$$

$$
\bar{R} = \begin{bmatrix} 9415.1 \\ 2077.5 \\ 2385.5 \\ 1843.1 \\ 2124.7 \\ 1883.7 \\ 1079.9 \\ 6308.5 \end{bmatrix}
$$

Step 3 (continued)

$$\bar{C} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \\ C_8 \end{bmatrix} \qquad \begin{aligned} \text{Where } C_i &= B_{1i} \times B_{2i} \times B_{3i} \ldots B_{8i} \\ \text{Therefore } C_1 &= 1 \times 4.2 \times 4.1 \ldots 4.8 \\ C_2 &= 4.0 \times 1 \times 3.1 \ldots 3.7 \end{aligned} \qquad \bar{C} = \begin{bmatrix} 19\,471.5 \\ 3\,282.7 \\ 2\,244.1 \\ 1\,909.4 \\ 778.8 \\ 1\,459.7 \\ 1\,266.4 \\ 5\,965.4 \end{bmatrix}$$

A final parameter T must be calculated in step four. T is the product of all the elements in R or C taken to the one-seventh power (0.1429). In order to calculate T this step is broken down into two operations.

Step 4

i)

$$\bar{R} = \begin{bmatrix} 9415.1 \\ 2077.5 \\ 2385.5 \\ 1843.1 \\ 2124.7 \\ 1883.7 \\ 1079.9 \\ 6308.5 \end{bmatrix} \qquad \bar{F} = \bar{R}^{0.1429} = \begin{bmatrix} 3.7 \\ 3.0 \\ 3.0 \\ 2.9 \\ 3.0 \\ 2.9 \\ 2.7 \\ 2.8 \end{bmatrix}$$

ii) T is the product of all the elements of F

$T = 3.7 \times 3.0 \times 3.0 \times 2.9 \times 3.0 \times 2.9 \times 2.7 \times 2.8$

$T = 8128.7$

Finally step five calculates the reduced diagonal element by using the following formula:

Step 5 $\qquad A_{ii} = (R_i \times C_i) / T$

Below are shown some examples as well as a table comparing adjusted and original diagonal values.

$$A_{ii} = (R_i \times C_i)/T$$

$$A_{11} = (9415.1 \times 19\,471.5)/8128.7 = 22\,681$$

$$A_{22} = (2077.5 \times 3\,282.7)/8128.7 = 843$$

$$A_{33} = (2385.5 \times 2\,244.1)/8128.7 = 662$$

Self citations	Adusted value	Original value
US	22 681	49 570
UK	843	4 875
WG	662	3 814
FR	434	2 953
SS	205	6 950
JA	340	3 835
CA	169	1 966
OT	4 654	12 038

Calculation of Domesticity and Internationality

It is interesting to note the difference between the original and adjusted diagonal. The adjusted diagonal is indicative of the self citations expected if each nation interacted with itself in exactly the same way it interacts with all other nations. The adjusted values are based on the number of times a nation cites and is cited, yet it is so much smaller than the original diagonal. The excess is caused by many factors but mainly it is due to one we may call domesticity. This is defined as the portion of a nation's papers that are only of interest to itself, dealing with domestic areas and purposes. These papers are by definition of no interest to other nations.

The difference between the original and adjusted diagonal is used to generate two parameters, domesticity and internationality. While domesticity is indicative of the portion of a nation's papers which circulate nationally, internationality is indicative of that portion of a nation's which circulate internationally. Formulas are given below.

$$\text{Domesticity} = \frac{\text{(Original diagonal value)} - \text{(Adjusted diagonal value)}}{\text{(Row sum using the Original diagonal value)}} \times 100\%$$

$$\text{Internationality} = 100\% - \text{Domesticity}$$

The chart below gives these values for Physics 1973.

Nation	Domesticity	Inter-nationality
US	38.1	61.9
UK	28.3	71.7
WG	23.8	76.2
FR	25.4	74.6
SS	46.2	53.8
JA	30.9	69.1
CA	25.6	74.4
OT	22.7	77.3

We can see from the table that Other Countries, West Germany, and France produced high percentages of their papers for international circulation for Physics 1973. In contrast the Soviet Union, the United States, and Japan had low values for internationality with correspondingly high values for domesticity. The most outstanding nation is the Soviet Union, which produces almost equal amounts of papers for national and international use.

Calculation of Input and Output Cofficients

Due to the structure of the citation matrix, input and output coefficients can be easily defined and calculated. The input coefficient is the number of times a nation is cited divided by total number of citations in the matrix multiplied by 100 percent. The output coefficient is the number of times that a nation cites divided by the total number of citations multiplied by 100 percent. A nation with a larger input coefficient than output coefficient is cited more than it cites, indicating that it exports more information than it imports. Conversely a nation with a larger output coefficient than input coefficient imports more information than it exports. The formulas of these parameters are given below.

$$\text{Input coeficient} = \frac{\text{(Column sum for a particular nation)}}{\text{Total citations in a given field}} \times 100\%$$

$$\text{Output coefficient} = \frac{\text{(Row sum for a particular nation)}}{\text{Total citations in a given field}} \times 100\%$$

Quality is defined as the input coefficient divided by the output coefficient. If this quantity is greater than unity then for this field the quality of science in a nation is above average. The nation's quality of science is below average if this parameter is less than unity. The formula for quality and a table showing input, output, and quality calculations for Physics 1973 is shown below.

$$\text{Quality} = \frac{\text{Input coefficient}}{\text{Output coefficient}}$$

Nation	Input size	Output size	Quality
US	50.8	37.1	1.37
UK	9.2	8.7	1.05
WG	6.7	8.6	0.77
FR	5.4	6.3	0.86
SS	2.8	6.7	0.42
JA	4.1	6.7	0.62
CA	3.9	4.5	0.87
OT	17.1	21.5	0.79

Input, output, quality, and internationality parameters were listed for all fields in 1973 and 1974. They appear in tables 2–3.

Expectation of Parameters

In order to determine if these input, output, quality, and internationality coefficients were reasonable the expectation of these parameters was calculated. This was accomplished by dividing each element by its column sum average. Most values are close to unity. Significant deviations indicate for a particular nation that its performance in science is above or below average for that field. Expectations are listed in Tables 4 and 5.

Examples of very high and very low values for input, output, and quality are given on Table 2. High values for input, quality, and internationality indicate a nation's strong contribution to science. This usually corresponds to a low output value. Low values for input, quality, and internationality indicate lesser contributions to science. This usually corresponds to high output values.

Table 2
Input and output sizes

$$\text{Input size} = \frac{\text{1973 Input Percentages}}{\text{1974 Input Percentages}}$$

	US	UK	WG	FR	SS	JA	CA	OT
Clinical medicine	$\frac{53.3}{54.5}$	$\frac{13.3}{13.3}$	$\frac{3.4}{3.5}$	$\frac{2.8}{2.6}$	$\frac{0.2}{0.2}$	$\frac{2.0}{2.3}$	$\frac{4.3}{4.1}$	$\frac{20.8}{19.4}$
Biomedical research	$\frac{54.5}{54.1}$	$\frac{11.6}{12.3}$	$\frac{5.2}{4.7}$	$\frac{3.6}{3.6}$	$\frac{0.7}{0.7}$	$\frac{2.9}{3.0}$	$\frac{4.1}{4.4}$	$\frac{17.6}{14.7}$
Biology	$\frac{46.1}{47.1}$	$\frac{14.4}{14.1}$	$\frac{4.2}{5.1}$	$\frac{2.4}{2.4}$	$\frac{0.4}{0.3}$	$\frac{3.7}{4.3}$	$\frac{10.1}{8.4}$	$\frac{18.9}{18.1}$
Chemistry	$\frac{42.0}{40.3}$	$\frac{11.8}{11.6}$	$\frac{6.4}{7.8}$	$\frac{5.0}{4.9}$	$\frac{2.5}{2.2}$	$\frac{6.2}{6.4}$	$\frac{6.7}{6.5}$	$\frac{19.4}{20.4}$
Physics	$\frac{50.8}{51.7}$	$\frac{9.2}{8.4}$	$\frac{6.7}{7.6}$	$\frac{5.4}{5.6}$	$\frac{2.8}{2.5}$	$\frac{4.1}{4.1}$	$\frac{3.9}{3.6}$	$\frac{17.1}{16.5}$
Earth & Space	$\frac{62.4}{63.3}$	$\frac{9.9}{9.0}$	$\frac{2.9}{3.7}$	$\frac{3.5}{3.4}$	$\frac{1.9}{1.3}$	$\frac{1.3}{1.3}$	$\frac{5.5}{4.5}$	$\frac{12.6}{13.6}$
Eng. & Tech.	$\frac{54.4}{51.8}$	$\frac{13.2}{13.1}$	$\frac{4.8}{5.7}$	$\frac{1.9}{2.5}$	$\frac{1.3}{1.1}$	$\frac{3.9}{4.4}$	$\frac{6.5}{6.9}$	$\frac{14.0}{14.6}$
Mathematics	$\frac{58.2}{54.0}$	$\frac{9.1}{8.6}$	$\frac{4.0}{4.6}$	$\frac{4.7}{5.6}$	$\frac{0.5}{0.3}$	$\frac{2.7}{2.3}$	$\frac{5.1}{6.0}$	$\frac{15.7}{18.4}$
Psychology	$\frac{76.0}{64.2}$	$\frac{6.7}{8.7}$	$\frac{0.5}{0.3}$	$\frac{0.9}{1.1}$	*	*	$\frac{8.9}{12.9}$	$\frac{6.7}{8.3}$

$$\text{Output size} = \frac{\text{1973 Output Percentages}}{\text{1974 Output Percentages}}$$

	US	UK	WG	FR	SS	JA	CA	OT
Clinical medicine	$\frac{42.8}{43.5}$	$\frac{9.4}{9.1}$	$\frac{7.6}{7.3}$	$\frac{6.1}{6.1}$	$\frac{0.9}{0.8}$	$\frac{2.9}{3.0}$	$\frac{4.5}{5.4}$	$\frac{25.9}{24.9}$
Biomedical research	$\frac{43.9}{45.6}$	$\frac{9.6}{9.9}$	$\frac{6.4}{6.5}$	$\frac{5.3}{5.3}$	$\frac{2.5}{2.1}$	$\frac{4.8}{4.7}$	$\frac{5.2}{5.0}$	$\frac{22.2}{20.9}$
Biology	$\frac{38.7}{39.3}$	$\frac{12.9}{13.2}$	$\frac{6.2}{6.8}$	$\frac{4.7}{4.5}$	$\frac{1.2}{1.1}$	$\frac{3.0}{3.1}$	$\frac{9.1}{9.2}$	$\frac{24.2}{22.8}$
Chemistry	$\frac{34.8}{34.1}$	$\frac{10.6}{19.0}$	$\frac{7.4}{6.9}$	$\frac{5.6}{4.6}$	$\frac{4.2}{2.7}$	$\frac{7.5}{6.7}$	$\frac{5.5}{5.0}$	$\frac{24.3}{20.9}$
Physics	$\frac{37.2}{38.3}$	$\frac{8.7}{8.3}$	$\frac{8.6}{9.4}$	$\frac{6.3}{6.2}$	$\frac{6.7}{5.8}$	$\frac{6.7}{6.3}$	$\frac{4.5}{4.2}$	$\frac{21.5}{21.5}$
Earth & Space	$\frac{51.9}{52.5}$	$\frac{10.3}{10.7}$	$\frac{4.3}{4.0}$	$\frac{4.3}{4.5}$	$\frac{2.9}{2.6}$	$\frac{2.8}{2.5}$	$\frac{5.4}{5.4}$	$\frac{18.3}{17.8}$
Eng. & Tech.	$\frac{43.7}{45.1}$	$\frac{10.0}{10.0}$	$\frac{6.9}{6.5}$	$\frac{3.7}{3.6}$	$\frac{2.6}{2.6}$	$\frac{7.3}{6.0}$	$\frac{6.1}{5.8}$	$\frac{19.8}{20.4}$
Mathematics	$\frac{47.1}{45.2}$	$\frac{8.2}{8.8}$	$\frac{7.6}{7.9}$	$\frac{4.5}{5.3}$	$\frac{1.6}{1.5}$	$\frac{3.6}{3.7}$	$\frac{6.4}{4.5}$	$\frac{21.1}{22.2}$
Psychology	$\frac{69.1}{51.1}$	$\frac{7.8}{10.1}$	$\frac{1.4}{2.4}$	$\frac{2.2}{3.0}$	*	*	$\frac{9.2}{17.2}$	$\frac{9.1}{14.6}$

Table 3
1973 and 1974 values for Quality

	US	UK	WG	FR	SS	JA	CA	OT
Clinical medicine	1.24 / 1.25	1.42 / 1.46	0.45 / 0.48	0.46 / 0.43	0.22 / 0.27	0.70 / 0.79	0.95 / 0.76	0.80 / 0.78
Biomedical research	1.24 / 1.19	1.21 / 1.24	0.80 / 0.72	0.68 / 0.68	0.26 / 0.32	0.59 / 0.64	0.79 / 0.88	0.79 / 0.83
Biology	1.19 / 1.20	1.21 / 1.07	0.80 / 0.75	0.68 / 0.54	0.26 / 0.31	0.59 / 1.40	0.79 / 0.91	0.79 / 0.80
Chemistry	1.21 / 1.18	1.11 / 0.61	0.86 / 1.12	0.89 / 1.06	0.59 / 0.81	0.83 / 0.95	1.22 / 1.30	0.80 / 0.97
Physics	1.37 / 1.35	1.05 / 1.02	0.77 / 0.80	0.86 / 0.91	0.42 / 0.43	0.62 / 0.64	0.87 / 0.85	0.80 / 0.77
Earth & Space	1.20 / 1.20	0.95 / 0.85	0.69 / 0.91	0.81 / 0.74	0.65 / 0.50	0.48 / 0.53	1.02 / 0.84	0.69 / 0.76
Eng. & Techn.	1.25 / 1.15	1.31 / 1.31	0.69 / 0.87	0.52 / 0.69	0.51 / 0.43	0.53 / 0.73	1.07 / 1.19	0.71 / 0.72
Mathematics	1.23 / 1.12	1.11 / 0.98	0.53 / 0.57	1.06 / 1.06	0.30 / 0.23	0.74 / 0.64	0.81 / 1.13	0.75 / 0.83
Psychology	1.10 / 1.34	0.86 / 0.86	0.36 / 0.13	0.41 / 0.37	*	*	0.93 / 0.75	0.74 / 0.57

Internationality 1973 and 1974

	US	UK	WG	FR	SS	JA	CA	OT
Clinical medicine	55.9 / 54.3	64.1 / 59.9	73.7 / 69.4	77.1 / 73.7	61.0 / 59.0	69.7 / 60.6	78.7 / 79.7	78.4 / 73.1
Biomedical research	61.8 / 59.4	71.5 / 67.6	74.0 / 73.2	68.3 / 67.8	64.0 / 58.3	74.1 / 68.1	78.5 / 72.3	78.8 / 73.6
Biology	47.9 / 43.1	56.6 / 50.8	67.5 / 61.9	61.9 / 57.2	75.0 / 67.7	38.7 / 31.3	59.6 / 55.2	72.1 / 64.9
Chemistry	64.4 / 61.7	68.1 / 73.9	59.8 / 52.0	53.9 / 47.8	39.8 / 29.3	57.5 / 51.3	65.9 / 63.5	64.5 / 56.5
Physics	61.9 / 57.9	71.7 / 68.7	76.2 / 72.8	74.6 / 70.2	53.8 / 51.6	69.1 / 62.5	74.1 / 73.0	77.3 / 73.6
Earth & Space	60.9 / 57.6	76.0 / 73.3	84.2 / 75.7	70.6 / 70.5	67.2 / 62.8	84.1 / 81.9	69.9 / 70.3	82.3 / 76.9
Eng. & Tech.	53.9 / 47.6	54.5 / 47.2	55.6 / 47.7	70.0 / 58.7	35.2 / 34.5	71.5 / 58.3	66.3 / 55.7	68.5 / 63.8
Mathematics	52.4 / 49.6	54.3 / 57.5	56.1 / 56.1	42.9 / 38.2	62.5 / 56.7	53.3 / 49.6	71.6 / 60.8	71.5 / 67.8
Psychology	*	*	*	*	*	*	*	*

*Not applicable.

Table 4
Expectation values for input and output coefficients
Expectation = (Element/Column sum avg) X 100%
(A) = Input expectation values for 1973 and 1974
(B) = Output expectation values for 1973 and 1974

(A)

	US	UK	WG	FR	SS	JA	CA	OT
Clinical medicine	0.96 1.02	1.21 1.21	0.80 0.73	0.83 0.74	0.16 0.19	0.60 0.65	0.70 0.64	1.31 1.21
Biomedical research	0.98 1.01	1.05 1.12	1.23 0.98	1.08 1.02	0.54 0.65	0.87 0.85	0.67 0.69	1.11 0.92
Biology	0.83 0.88	1.31 1.28	0.99 1.07	0.72 0.68	0.31 0.28	1.10 1.22	1.65 1.32	1.19 1.13
Chemistry	0.76 0.75	1.07 1.05	1.51 1.63	1.49 1.39	1.94 2.04	1.85 1.83	1.09 1.02	1.22 1.28
Physics	0.92 0.97	0.83 0.76	1.58 1.59	1.61 1.59	2.17 2.31	1.22 1.17	0.64 0.57	1.08 1.03
Earth & Space	1.13 1.18	0.90 0.82	0.69 0.77	1.04 0.97	1.48 1.20	0.39 0.37	0.90 0.71	0.79 0.85
Eng. & Techn.	0.98 0.97	1.20 1.19	1.13 1.19	0.57 0.71	1.01 1.02	1.16 1.25	1.06 1.08	0.88 0.91
Mathematics	1.05 1.01	0.83 0.78	0.94 0.96	0.40 1.59	0.39 0.29	0.81 0.65	0.83 0.44	0.99 1.15
Psychology	1.37 1.20	0.61 0.79	0.12 0.06	0.26 0.31	*	*	1.45 2.03	0.42 0.52

(B)

	US	UK	WG	FR	SS	JA	CA	OT
Clinical medicine	0.94 0.99	0.97 0.82	1.21 1.14	1.29 1.27	0.32 0.34	0.60 0.67	0.86 0.78	1.25 1.21
Biomedical research	0.97 1.04	0.99 0.90	1.03 1.01	1.12 1.11	0.88 0.88	0.99 1.04	0.99 0.72	1.07 1.01
Biology	0.85 0.90	1.33 1.20	1.00 1.06	0.99 0.94	0.42 0.46	0.62 0.69	1.74 1.32	1.17 1.10
Chemistry	0.77 0.78	1.09 1.73	1.19 1.07	1.18 0.96	1.49 1.13	1.55 1.49	1.05 0.72	1.17 1.01
Physics	0.82 0.87	0.89 0.75	1.38 1.47	1.33 1.29	2.37 2.42	1.39 1.40	0.86 0.60	1.04 1.04
Earth & Space	1.14 1.20	1.06 0.97	0.69 0.62	0.91 0.94	1.03 1.08	0.58 0.56	1.03 0.78	0.88 0.86
Eng. & Tech.	0.96 1.03	1.03 0.91	1.11 1.01	0.78 0.75	0.92 1.08	1.51 1.34	1.16 0.83	0.96 0.99
Mathematics	1.04 1.03	0.84 0.80	1.22 1.23	0.95 1.11	0.57 0.63	0.75 0.82	1.22 0.78	1.02 1.07
Psychology	1.52 1.17	0.80 0.22	0.22 0.37	0.46 0.63	*	*	1.83 2.47	0.44 0.71

Table 5
Expectation values for quality
Expectation = (Element/Column sum avg) X 100%

	US	UK	WG	FR	SS	JA	CA	OT
Clinical Medicine	1.02	1.26	0.70	0.67	0.54	0.99	0.98	1.05
	1.02	1.40	0.68	0.60	0.66	1.00	0.79	1.00
Biomedical Research	1.02	1.07	1.25	0.99	0.63	0.83	0.81	1.04
	0.98	1.19	1.02	0.94	0.78	0.81	0.92	1.06
Biology	0.98	0.99	1.06	0.74	0.76	1.70	1.13	1.02
	0.98	1.03	1.06	0.75	0.76	1.77	0.95	1.02
Chemistry	0.99	0.98	1.34	1.29	1.44	1.17	1.26	1.05
	0.97	0.59	1.59	1.47	1.98	1.20	1.35	1.24
Physics	1.12	0.93	1.20	1.25	1.02	0.87	0.90	1.05
	1.11	0.98	1.13	1.26	1.06	0.81	0.89	0.99
Earth. & Space	0.98	0.84	1.08	1.17	1.58	0.68	1.05	0.91
	0.98	0.82	1.29	1.03	1.22	0.67	0.89	0.99
Eng. & Tech.	1.02	1.16	1.08	0.75	1.24	0.75	1.10	0.93
	0.94	1.26	1.23	0.96	1.05	0.92	1.24	0.92
Mathematics	1.01	0.98	0.83	1.54	0.73	1.04	0.84	0.99
	0.92	0.94	0.81	1.47	0.56	0.81	1.18	1.06
Psychology	0.90	0.76	0.56	0.59	*	*	0.96	0.97
	1.11	0.83	0.18	0.51			0.78	0.73

Generating the Expectation Matrix

The input and output coefficients are used to generate a new citation matrix, which we define as the expectation matrix. This clearly shows if one nation is citing others more or less than would be expected on the basis of input and output parameters. The formation of this device can be better explained by calculating an element of it as an example. For Physics 1973, the number of times the United States cites West Germany is 2793. The input coefficient of West Germany is 6.55% (0.0655), while the output coefficient for the United States is 37.1% (0.371). Given that the total number of citations in the matrix is 117 421, we would expect the United States to cite West Germany (0.371 X 0.0655 X X 117 421 = 2897) times. The actual number is 2793. The expectation of this element is therefore (2793/2897) = 0.96). This means that the United States cites West Germany only 0.96 of the expected value based on the input and output coefficients of both countries. This element is placed in the expectation matrix for United States cites West Germany. The general formula is shown below:

Expectation for Country A citing Country B = (Input coefficient of B) × (Output coefficient of A) × × (Total number of citations in the matrix).

All other expectations for Physics 1973 were calculated and are shown below. Most values in the matrix are close to unity, however the ones that deviate significantly are indicative of which nations tend to work together for reasons such as common language or frequent interchange of researchers.

		Cited country							
		US	UK	WG	FR	SS	JA	CA	OT
Citing country	US	1.02	1.02	0.96	0.97	1.01	0.90	1.13	0.94
	UK	1.08	0.90	0.93	0.92	0.63	0.86	1.03	0.96
	WG	0.99	0.93	0.98	0.98	1.02	1.13	0.78	1.10
	FR	0.96	1.03	1.00	1.08	1.13	1.10	1.10	1.02
	SS	0.96	0.78	0.99	1.05	0.93	1.72	0.72	1.11
	JA	1.06	0.90	0.94	0.91	1.29	1.06	0.85	0.89
	CA	1.04	1.17	0.87	1.04	0.45	0.93	0.83	0.98
	OT	0.93	1.08	1.15	1.07	1.14	0.92	1.00	1.08

After computing the expectation matrices for all fields in 1973 and 1974, it was observed that all the matrices were very similar. Expectations between citation pairs were consistent. For example the expectation for France citing the US for all fields in 1973 and 1974 was consistently close to 0.96, which is the average of all the expectations for this citation pair. Therefore the average expectation matrix was calculated by adding all the matrices and dividing by 18 (for the nine fields in 1973 and 1974), and was used to approximate the flow of scientific information for all fields in for the two years. This would simplify the analysis and allow conclusions to be made. This matrix is shown below.

		(cited)							
		US	UK	WG	FR	SS	JA	CA	OT
(citing)	US	1.02	0.99	0.90	0.91	0.96	1.01	1.14	0.94
	UK	1.00	0.92	0.94	0.97	0.73	0.82	1.07	1.14
	WG	0.96	0.99	1.05	1.20	1.06	1.05	0.77	1.12
	FR	0.96	0.97	1.08	1.08	1.00	1.23	0.87	1.06
	SS	0.98	0.86	1.29	1.19	0.96	1.76	0.69	0.95
	JA	1.11	0.83	1.04	1.02	1.08	0.99	0.82	0.82
	CA	1.11	1.12	0.72	0.87	0.66	0.81	0.74	0.89
	OT	0.93	1.09	1.23	1.12	1.26	0.96	0.96	1.09

Table 6
Citing–cited relationships

Nation citing	Nation cited	Expectation
SS	JA	1.76
SS	WG	1.29
OT	SS	1.26
OT	WG	1.23
FR	JA	1.23
WG	FR	1.20
SS	FR	1.19
...	...	...
...	...	...
...	...	...
JA	UK	0.83
JA	OT	0.82
JA	CA	0.82
UK	JA	0.82
CA	JA	0.81
WG	CA	0.77
CA	CA	0.74
UK	SS	0.73
CA	WG	0.72
SS	CA	0.69
CA	SS	0.66

The expectation values for the 64 possible pairs of citing/cited nations form a Gaussian distribution of mean value 1.00 and standard deviation 0.17 showing that the simplest model of a total absence of international clustering will account for the data within an accuracy of 17%. Table 6 shows the values of expectation that lie more than one standard deviation above and below the norm. It would seem that the only singularities that can be attributed with some certainty to effects beyond the range of random noise are the unexpectedly strong interaction with which the USSR cites Japanese science, and the mutual avoidance that seems to exist between Canada and the USSR. It is worthy of note that one does not see strong clustering based on linguistic similarities and differences, these seem to exist but only weakly, below the level of a single standard deviation. A more detailed analysis of expectations for each of the separate fields of science yields similar data, though there seem to be other idiosyncracies that look real enough since they persist from the 1973 data to that of 1974. Further analysis of larger data banks for more years of data and perhaps fuller disaggregation seem to be very much worthwhile.

In addition to the structural clustering, or rather, the lack of much of it, some nations in some fields show parameters of input, output or quality that are

Table 7
High expectation values for input, output, and quality in years 1973/1974

	Field	Country	Value
Input	PH	SS	2.17/2.31
	CH	SS	1.94/2.04
	CH	JA	1.85/1.83
Output	PH	SS	2.37/2.42
	CH	JA	1.55/1.49
	PS	CA	1.83/2.47
Quality	MA	FR	1.54/1.47
	BI	JA	1.70/1.77
	CH	SS	1.44/1.98

unexpectedly high. They are shown in Table 7 and agree rather well with intuitive judgment. Again what is surprising is that there are very few exceptions and even those do not rise far above the level of random fluctuations from the most simple model. It therefore seems that we have a parsimonious method for reducing large blocks of data to few parameters. The method here has been used for data disaggregated into a small group of the largest nations and only the few chief broad categories of science. A larger set of matrices for more countries and narrower categories of science could be tackled in the same fashion and we are confident it would yield similar but more precise results. The citation behavior of nations is governed by their size, quality, and internationality with very little fine tuning for clustering structure between nations and only weak fine tuning rather than strong differences from field to field for any nation. The amount of regularity thereby uncovered is a little surprising; there is not much room for the making of special allowances for language clusters, special national interests, or special subjects of high quality in one nation or another.

In addition to this investigation by nation and field, the same technique can, we feel sure, be used to comparable effect in the analysis of the large scientometric data bank provided by the *Journal Citation Reports* which gives the citations from each ISI source journal to every other journal. Quite clearly each journal has a size and a quality. We now suppose, as has not hitherto been done, that each journal will have a coefficient of "domesticity" governing its self-citing behavior which is habitually far in excess of that expected if each journal treated itself as it treats others. Over and above this it seems very likely that journals must cluster much more decisively than do nations, giving us minor empires of mathematics, physics, etc. Whether this is so is another topic for further investigation.

Notes and References

1. We thank the Institute for Scientific Information and its president Eugene *Garfield* and research chief Henry *Small* for much help with this and other citation information extending over many years of the Institute's contribution to the information sciences.
2. We would like to thank Francis *Narin* for providing this material in the form of printouts prepared for the National Science Board. The long and arduous task of analysing this data was designed and executed by Mark *Carpenter,* and the work for the *Science Indicators* was organized by Robert *Wright;* both are hereby thanked for this valuable contribution.
3. *Science Indicators 1974,* National Science Board, Washington, DC, GPO 038-000-00253-8, 1974, 157–161.
4. D. de S. PRICE, "The Analysis of Scientometric Matrices for Policy Implications", *Scientometrics,* 3 (1981) No. 1, 47–54; and The Analysis of Square Matrices of Scientometric Transactions, *Scientometrics* 3 (1981) No. 1, 55–63.
5. We thank Worcester Polytechnic Institute for the use of their Computer Center. Thanks also are extended to the many people who contributed at the Institute, including faculty advisors Dr. Arthur *Gerstenfeld* and Dr. Patrick *Dunn,* professors in the Mathematics Department and the students at the Computer Center for their help.
6. M. CARPENTER, *International Science Indicators–Development of Indicators of International Scientific Activity Using the Science Citation Index,* 1979 CHI Computer Horizons, Inc., 1050 Kings Highway North, Cherry Hill, NJ.