Scientometrics, Vol. 1. No. 1 (1978) 43–64

WORLD SCIENCE AS AN INPUT-OUTPUT SYSTEM

H. INHABER,* M. ALVO

Planning and Finance Service, Environment Canada,
Ottawa, K1A OH3 (Canada)

(Received March 20, in revised form May 4, 1978)

World science can be characterized as the product of one scientist or nation – knowledge or published papers – used or consumed by other scientists or nations. In this sense, science can be viewed as an input-output system, analogous to the models used in economics. An input-output model of the citation patters of the 18 leading countries in international science was constructed. These countries produce most of the world's science. The large role of the United States in both producing and consuming scientific information is evident in the results. The models also show the role of other countries with respect to each other. For example, the multinational nature of science in countries like the Netherlands and Switzerland is evident. The model can be used to show which countries interact with others, and which do not. Both types of information are useful in discussing trans-national interactions in science.

Introduction

KNAPTON[1] has written of FRANCOIS QUESNAY, the originator of the concept of the input-output table in 1758, that "One of Quesnay's most ardent contemporary admirers ranked the *Tableau Economique* with the invention of writing and the invention of money as the three greatest discoveries since the world began". Even if, allowing for some overstatement, the input-output (I/0) concept ranked in the first hundred rather than the first three most important human inventions, it is not clear how it can be related to the publication and information network of world science. It is the object of this paper to demonstrate that this network is similar in many ways to I/0 tables, and to draw conclusions unobtainable by other methods.

Input-output outside economics

I/0 tables can be defined as measurements of how different industries in a country or region interact with each other. This interaction is usually measured in terms of money or goods but the concept can be extended far beyond the economic realm.

*Present address: Atomic Energy Control Board, Ottawa, KIP 5S9, Canada.

For example, I/0 has been used in discussions of environment and pollution questions,[2–16] agriculture,[17–20] energy,[21–23] air traffic,[24] education,[25] hospital planning,[26] wages[27] and administrative processes.[28]

Clearly, with this wide variety of uses, I/0 may be applied to other fields. Because I/0 is simply a set of mathematical techniques, it can, in principle, be devoted to any problem which is comprised of a group of interactions.

World science as an I/O problem

World science can be seen as operating in an I/0 framework. In this view, nations take the place of industries in the economic I/0 system. Country A produces a given quantity of science, some important and some less so. Country B uses part of it, as well as producing its own. In turn, country A uses part of B's contribution. Expand this to enough countries, and one has a model of world science interaction.

This is, of course, a highly simplified model. Science, if it is to be divided at all, is usually broken down by subject, although even here boundary lines are notoriously imprecise. However, in this era of national science policies, it is of interest to know how the science of one nation interacts with that of another. In addition, knowledge of these factors would tell us the effect of language, economic, political and other barriers to the spread of knowledge.

How can we define the interactions of world science in operational terms? It is well known that much of this interaction is informal, consisting of conversations and letters. While attempts have been made to model this for small communities of scientists, it would not be feasible for larger groups. On a global basis, the only simple way of modelling interactions is by means of formal communications by scientific articles.[29] The field is extensively reviewed by NARIN.[30]

A journal article is only one direction on the two-way street of interaction. It tells us what has been done, but not how many scientists read it or acted on it. In principle, this quantity is probably unknowable; in practice, we approximate it by means of references to the paper by other papers.

The hazards of this citation analysis are clear; NARIN[30] lists most. However, this statistical method allows us to gain insights impossible by other analytic procedures.

In this paper, we shall determine the national origin of scientific articles, and in turn determine the origin of the articles which cite them. For example, paper X from country A may be cited 10 times by scientists from country B, 20 times from country C, and so on. The interactions of the 18 largest scientific nations, which produce over 95% of the world's major scientific literature as computed by the *Science Citation Index*,[31] will be discussed. Because the number of possible national interactions is huge, a technique for illustrating those of most interest will be used.

Previous work

Perhaps the first effort to evaluate scientific interactions using citations is attributable to CASON and LUBOTSKY.[32] They measured the influence (or interaction) of 28 psychological journals by constructing a matrix of the journal citation rates. Although not labelled an input-output table, the scheme clearly contained the germ of the idea. Because four decades have passed since their work, we will not present their results here.

Evaluating the interactions of many journals is a time-consuming task. A simpler approach, taken by some, is to consider the interactions of one journal or nation with all the rest. This was the system adopted by GUPTA,[33] who considered the citation patterns of Indian physicists. In effect, this analysis considers only one row (or column) of the potentially large input-output matrix.

GUPTA found that about 60% of all citations made from a prominent Indian physics journal were to publications originating in the U.S. or the U.K., with only 14% to Indian publications. While instructive, the work has at least three defects: (a) only one journal was considered, (b) citations to, as opposed to references from, Indian work were not considered, and (c) articles published in a particular country were assigned to that country, regardless of the author's national origin. Nonetheless, GUPTA was one of the pioneers in this field.

Because of the sheer mass of citations and references in the scientific literature, studies of international interactions have tended to concentrate on one nation or branch of science. This limitation was overcome in the past decade by the annual publication of the *Science Citation Index* and the availability of the computer tapes from which it was compiled. These tapes were used by Computer Horizons Inc. of Cherry Hill, N. J. to produce a series of articles on national aspects of world science.

In their first publication,[34] Computer Horizons analysed citation patterns between six major scientific countries: United States, the U.S.S.R., the U.K., Japan, Germany and France. The study covered all of science, and in addition subdivided it into fields like physics, chemistry, mathematics and engineering.

It would be impossible to summarize all the results of this paper. However, it was found that the U.S. was by far the most highly cited country, followed by the U.K., Germany and Japan. French and Soviet publications were the least heavily cited. The U.S. held the lead in each of the seven fields considered, even when the number of publications was taken into account, except for systematic biology. In this field, Germany had the greatest proportion of citations to its number of papers.

While the Computer Horizons paper marked an advance over previous work, there were still deficiencies. First, papers in a journal were assigned, for citation purposes, to the country in which that journal was published, regardless of where the paper actually originated. For example, a Canadian paper appearing in *Physical Review* is assigned to the U.S. Second, only six countries were considered. While it is clearly impractical to consider all the scores of countries which now produce more than a few scientific papers each year, more than a small number are needed to visualize the main interactions. Finally, there was no theoretical framework into which the data fitted. Little discussion was given on the revealed interaction patterns or which figures were more significant than others. The objective of the present paper is to fill these gaps.

Input-output

Before proceeding further, a few words on the I/0 concept are necessary. As mentioned above, its prime use is in economic theory, although it has been applied to other fields. Suppose we have an economy made up of three sectors: farming, machinery and food processing. Farmers sell their crops to the food processing industry. Machinery is sold to farmers and food processors. Food processors sell their foods to the machinery industry and farmers. As well, there are transactions within each industry.

A hypothetical case is shown in Table 1, using arbitrary units. Food processors bought 50 units from farmers, and farmers in turn bought 6 units from food processors. The former was the largest transaction and is thus the most important component in the table. In general, the matrix is not symmetric around the diagonal.

In this brief discussion, we have ignored the finer points of economic analysis, such as the "final demand" and "value added" matrices. In addition, the numbers of Table 1 include services as well as goods, and comprise intermediate goods as well as final goods. Leaving these details aside, Table 1 does show the major interaction patterns of an economic system.

We can use the same reasoning on international science. If we substitute nations for the economic sectors of Table 1, and "citations to" and "references from" for

Table 1
Hypothetical I/0 table for a three-compnent economy

		Sold to		
		Farmers	Food Processors	Machines
Bought from	Farmers	10	50	0
	Food Processors	6	5	30
	Machines	30	10	40

the goods and services bought and sold, we have a simple model of how science interacts across national boundaries. Suppose country A never cites country B, and in turn is never referred to by country B. Although there may be some informal interaction between these countries, it is fair to assume that overall interaction is slight at best. There will clearly be exceptions to this, especially in the case of political interference, but they will probably be few in number.

Conversely, large matrix values connecting countries likely indicate strong interactions. Of course, any science input-output matrix must be corrected for the widely varying sizes of the scientific enterprise from country to country.

This model does not account for all aspects of scientific interaction. Much occurs within national boundaries, as we shall see. However, even the degree of this insularity gives us a clue as to how inward-looking a nation is.

Put briefly, we shall use I/0 as a model for international science. The size of the matrix elements tells us the degree of interaction.

Mathematical implications of I/0

While the I/0 matrix shows relationships clearly when it is small, once it is of any magnitude we become lost in a welter of numbers. What is needed is a simple graphical portrayal of the main relationships as well as the matrix numbers.

This problem has been considered before. XHIGNESSE and OSGOOD[35] used a graphic portrayal of the "distance" between psychological journals in terms of their citation patterns. Their work can be traced to that of DANIEL and LOUTTIT,[36] who also worked with psychology journals.

Perhaps the simplest method of determining how closely the vectors – rows or columns – of a matrix are related to each other is the multidimensional scaling system described by KRUSKAL.[37,38] Without going into the detailed mathematics, the scaling calculates a "stress", which is a measure of the mathematical distance between vectors. Subject to certain constraints, the stress is minimized by an iteration process until it reaches its lowest possible value.

The advantage of this method in terms of understanding is that the stresses can be considered as being reduced to two dimensions, and so can be plotted. The points for each nation indicate the relative position each occupies with respect to an arbitrary origin. If, for example, nations A and B are close in this plot, we know that their citation patterns are similar.

In addition, we can perform simple statistical tests to determine why two countries are similar (close) or dissimilar (far apart) in their patterns. For example, countries A and B may be similar mainly because they both cite country C almost exclusively, and not because of their relationship to country D.

In summary, there are at least three methods of presentation. First, the bare I/O matrix; second, multidimensional scaling plotted graphically; and third, statistical tests on particular matrix elements.

Location of scientific publishing

If all scientists published in their own countries, the problem of tracing international interactions would be made simpler. However, this is often not the case. A scientist from country X may publish in country Y's journal for a variety of reasons – political, greater publicity, prestige, or simply because he happens to be there temporarily. In any case, we cannot consider scientific interaction unless we allow for publications crossing national borders.

One of the first analyses of this phenomenon is due to INHABER.[39] Using a sample of publications from the *Science Citation Index,* he found that publishing patterns varied sharply among nations. For example, authors from the U.S. wrote about 77% of that nation's publications. Authors from the Netherlands, however, made up only about 5% of Dutch publications.

The INHABER paper was based on sampling, and so could contain statistical errors, particularly for the smaller countries. A comprehensive study of the 278000 papers noted in the 1973 edition of the *Science Citation Index* is available.[40] This matrix, shown in Table 2, indicates the fraction of authors who published *in* country L *from* country M. For example, 45% of articles published in Britain are from British authors, 21% from the U.S., and 4% from Canada, reading the vertical columns. The vertical column will add to 1 if all nations are considered; Table 2 shows only some major nations.

Table 2
Partial matrix of where scientists publish (after NARIN[40])

	Publishing in							
Authors from	U.S.	U.K.	Germany	Netherlands	U.S.S.R.	Canada	Japan	France
U.S.	0.751	0.211	0.097	0.346	0.001	0.174	0.032	0.020
U.K.	0.037	0.450	0.046	0.108	0	0.015	0.003	0.009
Germany	0.015	0.023	0.531	0.059	0.002	0.002	0.002	0.009
Netherlands	0.012	0.006	0.009	0.052	0	0	0	0
U.S.S.R.	0.004	0.012	0.005	0.088	0.987	0.001	0.001	0.004
Canada	0.045	0.040	0.015	0.054	0	0.755	0.006	0.009
Japan	0.025	0.020	0.020	0.087	0	0.005	0.910	0.002
France	0.017	0.022	0.032	0.032	0	0.011	0.001	0.834

Reading horizontally, the numbers, indicate the relative proportions of authors from a particular country publishing in another. The horizontal fractions do not add to 1, since each fraction refers to a different sum. For example, 4400, or 3.7% of the 1973 articles in U.S. publications are British in origin; 950, or 4.6% of articles in German publications are also British; and so on.

We will use this information to "deconvolute" the citation data obtained.

Methodology

The data base for this study was the *Journal Citation Report* (published by the Institute for Scientific Information, Philadelphia). The 1973 edition, applicable to the last quarter of 1969, was used. While using data for only one quarter may distort information for a few small journals, it is unlikely to do so for entire nations.

A sample of the data structure is shown in Table 3. The *Journal of American Chemical Society (J.A.C.S.)* made 19,620 citations during the period considered. Of those, 3503 were to itself, 767 to the *Journal of the Chemical Society, (J.C.S.)* and so on. All journals which received three or more citations from a particular journal were tabulated. In addition, the citations were shown by the year of the article being cited. For example, *J.A.C.S.* made 52 references to 1968 articles in the *Journal of Chemical Physics (J.C.P.).*

The first step in assigning these citations on a national basis is to identify the origin of each journal. We then add the citations assigned to each country. In the example of Table 3, 19620 references are made from the U.S. *(J.A.C.S.),* of which 3975 *(J.A.C.S.* and *J.C.P.)* are to the U.S., and 767 *(J.C.S.)* are to the U.K.

The are a large number of sources to which citations can be made. In the example being discussed, *J.A.C.S.* referred to 228 sources three or more times. To make the data more manageable, a cut-off point of 10 or 15 citations was used.

The data were then accumulated by country, and a partial matrix is shown in Table 4. As an example, U.S. publications referred to Soviet publications 4650

Table 3
Sample data from Journal Citation Report, 1973

	Total	1969	1968	1967
Citing journal:				
J. Am. Chem. Soc.	19620	1538	3029	2464
Cited journals:				
J. Am. Chem. Soc.	3503	399	628	436
J. Chem. Soc.	767	60	153	99
J. Chem. Phys.	472	20	52	42

Table 4
International citation input-output matrix (15 citations or more)*

Citations made	Citations received								
	U.S.	U.K.	Germany	Netherlands	U.S.S.R.	Canada	Japan	France	India
U.S.	164 300	23 600	4 660	7300	930	360	330	810	72
U.K.	21 300	20 900	550	1100	89	250	210	160	–
Germany	15 100	3 790	12 400	1100	380	129	270	310	–
Netherlands	15 000	2 880	820	7500	300	180	210	54	–
U.S.S.R.	4 650	670	310	620	15 700	42	94	21	–
Canada	3 210	1 110	160	350	–	1200	39	40	–
Japan	5 220	1 140	450	511	120	45	2800	53	–
France	4 930	1 400	600	180	110	52	55	4600	–
India	780	390	58	–	–	–	–	–	520

*This means, for example, that U.S. publications made about 930 citations to U.S.S.R. publications in the period under consideration. Conversely, Soviet publications made 4650 citations to U.S. publications.

times in the period under consideration. The number of citations to U.S. publications from Soviet publications was 930.

In order to save space, Table 4 shows only nine of the 18 countries considered. Nationals of the 18 countries produce about 92% of the world's scientific literature. Publications of the 18 countries produce slightly more, so that the matrices we consider comprise almost the complete set.

In Table 4, the cut-off point of 15 citations was used. This accounts for some of the blanks. For example, Indian publications may have cited Canadian journals, but not one Canadian journal 15 or more times.

Using different cut-off points and dates of citations, as well as counting the number of journals cited, a total of seven I/0 matrices were prepared. However, we have pointed out previously that the place of publication does not always correspond to the author's origin. To circumvent this problem, each of the seven matrices was multiplied by the matrix represented in Table 2. This automatically assigns the papers published in a particular country to the correct proportion of authors from all countries.

Our major assumption in this procedure is that citations to papers in a journal are not biased with respect to their national origin. For example, suppose a journal has 10% of its authors from country A, 20% from B, and 70% from C. We assume that citations to papers in that journal will be in the same proportion. While this assumption clearly will not hold for a particular journal, the large number of papers and citations being considered probably make it valid in the whole. This does not imply that citations by a particular scientist will not tend to favor his home country's work.

Graphical results

Input-output is a useful model for international science interactions. The following figures and mathematical results illustrate its significance. Because of the large number of possible interactions which can be considered, only the highlights are shown here.

Fig. 1 shows the publishing relationships between 18 countries, as deduced from the complete version of Table 2. The data was multi-dimensionally scaled, as described in the Appendix.

In general, the deductions which can be made from an inspection of Table 2 are borne out in Fig. 1. For example, a country like the Netherlands has only about 5% of its papers produced by its natives, an anomalous situation compared to that of other countries. Countries with this large degree of "multi-nationality" in publishing behavior within their own country are grouped to the bottom left-hand corner

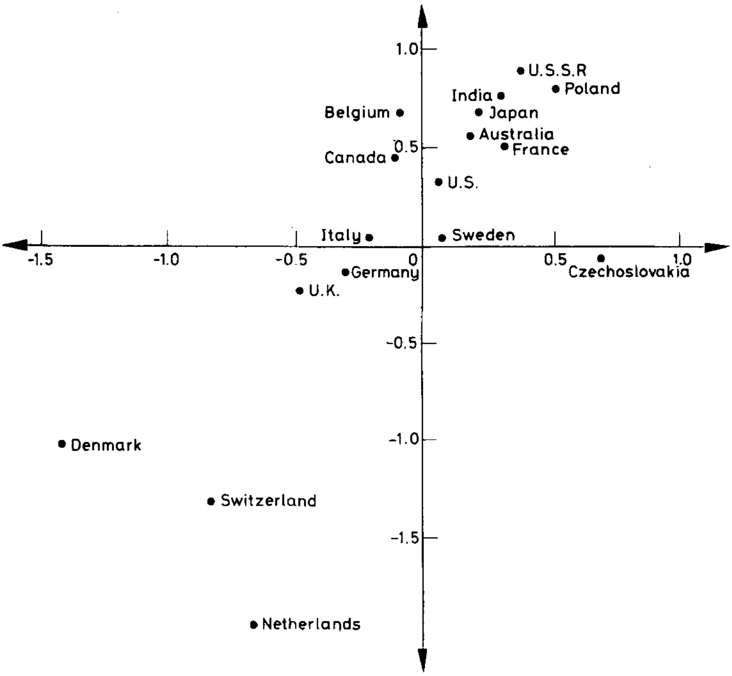

Fig. 1. Relationship of nations by publishing patterns (using multidimensional scaling). Based on Table 2, this array shows clearly that "multi-national" publishing countries like Switzerland, the Netherlands and Denmark have patterns very different from that of other nations. As in subsequent figures, the absolute values of the scales mean little; only the relative positions are significant

of Fig. 1. Those nations whose publishing effort is confined mostly to their own nationals are mostly in the upper right-hand corner. For example, the U.S.S.R. has 98.7% of its output from residents of that country. Corresponding values for India are 95.0%; for Poland, 92.3%. Nations like the U.S. and the U.K. occupy an intermediate position.

Fig. 1 shows clearly the degree of publishing insularity. Denmark, Switzerland and the Netherlands are clearly at the opposite end of the scale from countries which publish almost exclusively the work of their own nationals.

As mentioned above, we can take account of the fact that publications in a given country are not always those of natives. We do this mathematically by multiplying the transpose of Table 4 by Table 2 (from the left). The transpose is needed to preserve the relationship between countries. Results are shown in Fig. 2.

The U.S. location is far from that of other countries. However, now we can see more of a pattern emerging among the other nations. Germany and the U.K., with

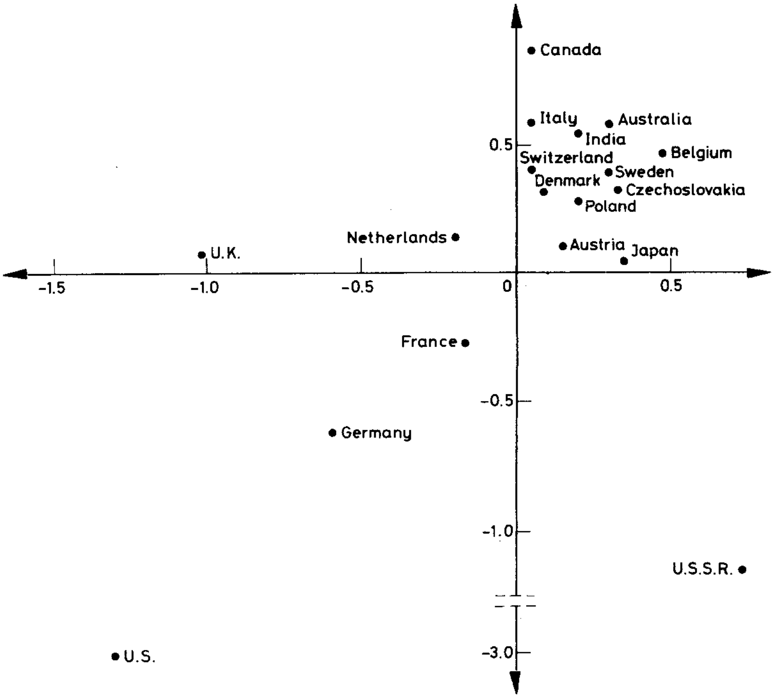

Fig. 2. International citation patterns, corrected for publishing patterns.Data has been modified by noting that papers published in a given country do not always originate from the nationals of that country. The location of the U.S. is far from other countries

their relatively strong dependence on U.S. science, are the countries closest to it. As an example, Table 4 shows that the U.K. has more citations to U.S. science than it does to British papers. The proportion of German citations to U.S. science to that of German papers is also about one. As well, U.S. science cites papers from the U.K. and Germany fairly stongly.

Why then is Canada so far removed from the U.S. in Fig. 2 ? The answer can be derived from a comparison of the first row and column of Table 4, although the results should be modified by the appropriate multiplication of Table 2. We see that Canada refers to US. publications substantially. In fact, they constitute almost three times the number of those to Canadian publications. However, the U.S. refers only infrequently to Canadian publications – a total of only 360 times, or about 0.2%. In other words, the distribution of citations to and references from is highly assymetrical between the two countries. This produces a large distance between them in Fig. 2. The same reasoning can be applied to other pairs of countries.

While a discussion of which countries are significantly close to (or apart from) others will be held in the next section, the relationship of the U.S.S.R. to the U.S. in Fig. 2 deserves comment. Their relative closeness arises from the fact that papers in the U.S.S.R. cite U.S. papers frequently. The ratio of Soviet citations to U.S. papers to that of their own papers is about one-third. Almost all papers published in the U.S.S.R. are by Soviet nationals. The ratio of Soviet to U.S./U.S. to Soviet citations is about 5 in Table 4. The corresponding ratio for Canada and the U.S. is about 10, implying less similarity of citation patterns. There is thus a link between Soviet and U.S. positions in Fig. 2, although not a strong one.

So far we have considered citations made in 1969 to all previous papers. But the weight of history can strongly affect the results. For example, German science dominated the world for decades. It may well be that most of the citations to papers originating in Germany are to those which were published long ago. Conversely, a country which is relatively new on the scientific scene, like India, will have most of the citations it receives to its papers of recent origin. The Institute for Scientific Information data allows us to subdivide the citations by years. A study of time distribution of citations has already been performed.[41]

Fig. 3 shows the multi-dimensional scaled distribution for citations to papers published within two years of the citing paper. In analogy to Fig. 2, the scaling was done on the matrix corresponding to Table 2 multiplying the transpose of a matrix similar to Table 4. This latter matrix contained only those citations which were made to recent papers, as defined above. Only those journals which cited another 15 or more times were considered in the analysis.

The overall pattern of Fig. 2 is retained, i.e., the U.S. is far removed from all other countries because of the large number of citations it receives. However, there are differences. For example, Canada is now much closer to the U.S. position than in Fig. 2. This is mainly due to the number of citations to U.S. journals from Canadian journals. They numbered about 50% more than the citations made within Canada.

Conversely, the U.S.S.R. is now far away from the U.S. Soviet citations to U.S. journals constituted only about 8% of all references from the former country. Of those citations to Soviet journals, less than 5% come from the U.S. Clearly there is little interaction as far as new work is concerned.

The question of language plays a large part in the citation pattern to new work. Many Soviet publications are translated into English for the benefit of U.S. readers, and this imposes a time lag. As a result, citations to Soviet work often have a long delay, and so this work may not be cited promptly. On the other hand, Canadian papers are usually in English, making communication with the large science base in the U.S. both quick and easy. This may also account for the relative closeness of

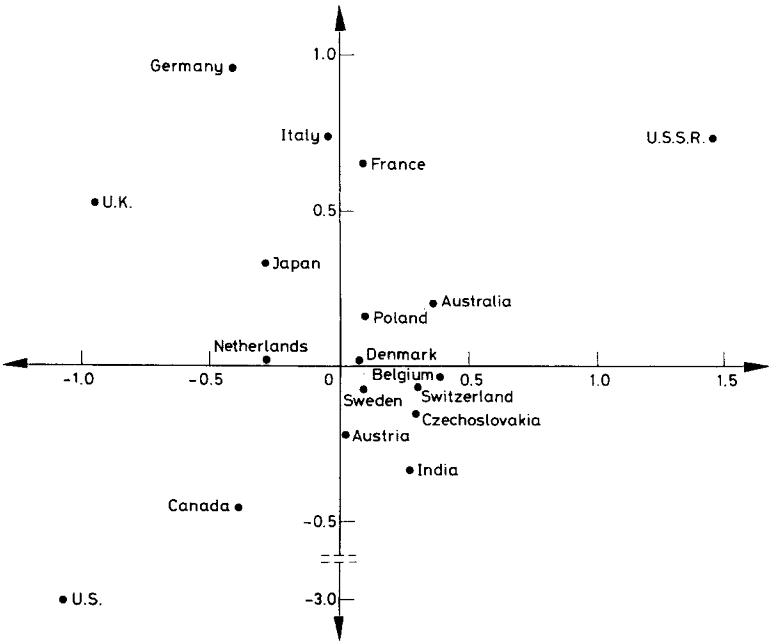

Fig. 3. Patterns of up-to-date international citations. By "up-to-date" we mean citations made to papers published within two years of the citing paper. The distribution by country, as shown by the usual multi-dimensional scaling, varies from the pattern produced when all citations are considered

India to the U.S., though not that of the U.K. or Australia. Other factors, such as the number of "letter"-type journals in a country, may affect the results.

Until now, we have considered only those journals which received 15 or more citations in the period under consideration. In principle, we should study the entire number of citations, down to one citation to one journal from another. In practise, the volume of citations makes this impossible. The ISI uses a cut-off of three citations, for example.

For countries with small scientific establishments, it may be useful to extend the cut-off point below 15 citations, since their absolute rate of citations and references will tend to be low. An illustration of the changes which occur when a cut-off of 10 citations is used is shown in Fig. 4. As in the two preceding figures, the calculations were made by multidimensionally scaling the product of (a) the extended Table 2 with (b) the transpose of the matrix (analogous to Table 4) comprising all citations made to journals receiving 10 or more citations.

We can then compare Fig. 4 with Fig. 2. The U.S. is still far from other nations, but the distance has been reduced considerably. The general layout of the two dis-

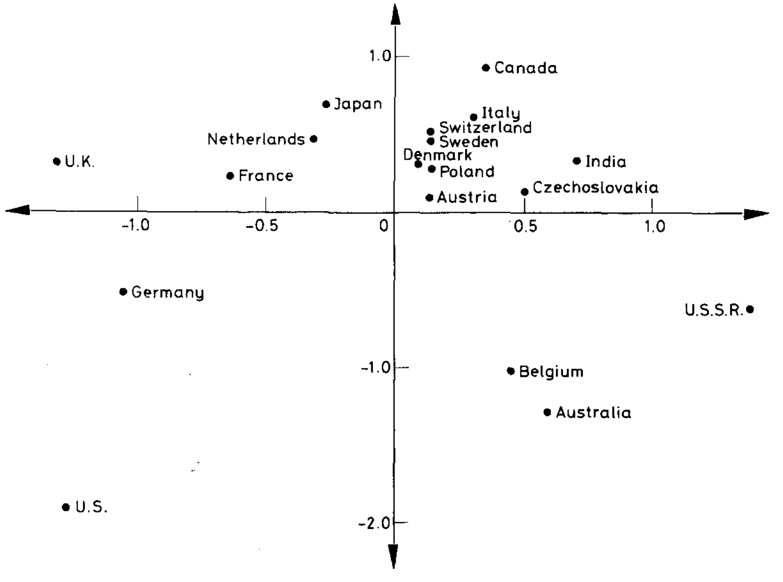

Fig. 4. International citation patterns for journals receiving 10 or more citations

tributions is similar, but there are some exceptions. For example, Belgium and Australia are now much closer to the U.S. than before. This indicates that journals originating from these two countries both make citations to and receive references from U.S. journals on a small scale, a scale which is overlooked in Fig. 2.

Finally, we can make a graphical allowance for the fact that some countries have many more journals – and thus more citations to these journals – than other countries. To do this, the number of journals which were used to compile Table 4 were displayed in a similar matrix from. Then each of the elements of Table 4 were divided by the corresponding element from the "journal" matrix. Where both elements were blank, the resulting element was assigned a value of zero.

Fig. 5 shows the results of this division. One of the major differences between this distribution and previous ones is that the U.S. is now much closer to other countries. In effect, the consequences of the large number of citations it receives have been removed. The U.S. cited about 3400 journals in the network and received references from about 3800. A second consequence is a clustering around the U.S. of three countries: the U.K., Germany and the Netherlands. The first two are well-known associates of the U.S. in scientific enterprises. The location of the Netherlands is due to a large proportion of its publications (about 35%) originating from U.S. authors. Many of these U.S. authors obviously cite Dutch publications, and vice versa.

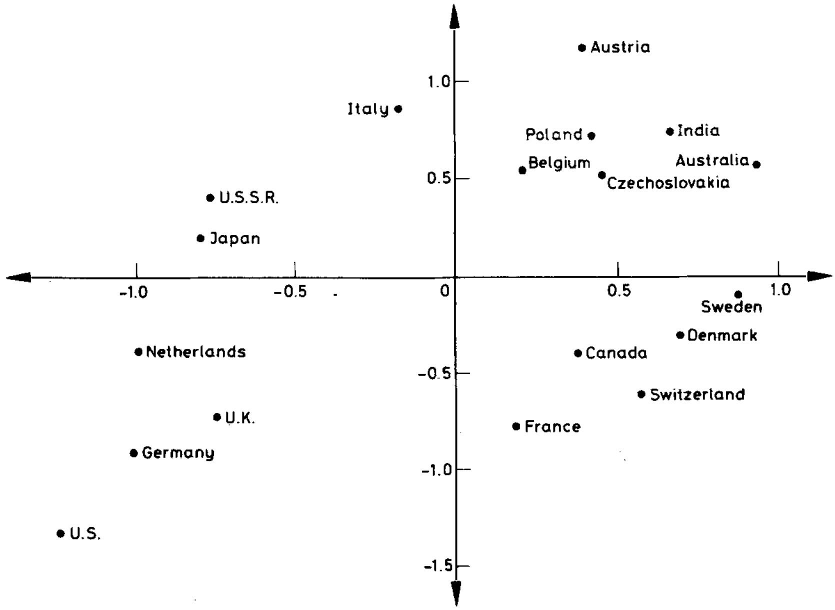

Fig. 5. Citation patterns, allowing for numbers of journals. In this way, the fact that the U.S. has such a high proportion of all citations is somewhat mitigated mathematically. The resulting distribution differs from previous examples.

Another point of interest is the relatively large distances between the U.S.S.R. and the two Eastern European countries in the group, Poland and Czechoslovakia. Since there is extensive cooperation between these countries, one might expect greater interaction. However, in the period under consideration Soviet journals made no citations to Polish or Czech journals. Polish and Czech journals made 0 and 46 citations to Soviet journals, respectively.

The latter value constituted only about 1.2% of all non-Soviet citations to Soviet journals. This data accounts for the large distances between the U.S.S.R., Czechoslovakia and Poland. On the other hand, U.S. citations to Soviet journals constituted about 59% of all non-Soviet citations to these journals. This fact brings the U.S.S.R. fairly close to the U.S., although not as close as the U.K. and Germany.

Figs 1–5 contain a large amount of information on inter-relationships between countries. Only the highlights of each figure have been mentioned. For those interested in a particular nation or group of nations, deeper analysis can be undertaken.

Statistical tests

As an example of this analysis, we can perform various statistical tests on the relationship between pairs of countries. The mathematical background is discussed in the Appendix.

The problem to be considered may be put into simple terms: how close is close? If we see Country A next to Country B in one of the previous figures, is it a mere coincidence, or is there a statistically significant relationship between them? The same question applies if they are greatly separated.

Too much space would be required for a complete listing of all the possible two-country relationships. We will consider only those few which are statistically significant.

Table 5 shows relationships between the U.S., the U.K., the U.S.S.R. and Canada for data uncorrected for publishing patterns, i.e., the fact that authors do not always publish in their home country. Potential international relationships not shown are not statistically significant.

We see that self-citation is high for the U.S., with this occurring with greater frequency than for the U.K., the U.S.S.R. and Canada. This is obviously reflective of the scientific strength of the U.S. A further indication of this is the fact that the U.K., the U.S.S.R. and Canada also cite the U.S. more than the U.S. cites them.

Further data is shown in Tables 6 and 7, corresponding to Figs 2 and 3, respectively. The mathematical assumptions for each set of data have already been discus-

Table 5
Analysis of citation patterns for journals with 15 or more citations, uncorrected for publishing patterns

U.S. – U.K.:	(a)	U.S. cites itself more than the U.K. cites itself
	(b)	U.K. cites the U.S. more than the U.S. cites the U.K.
	(c)	U.S. cites the U.S.S.R. more than the U.K. does
U.S. – Canada:	(a)	U.S. cites itself more than Canada cites itself
	(b)	Canada cites the U.S. more than the U.S. cites Canada
	(c)	Canada cites the U.K. more than the U.S. does
	(d)	U.S. cites the U.S.S.R. and Italy more than Canada does
U.K. – Canada:	(a)	U.K. cites itself more than Canada cites itself
	(b)	Canada cites the U.K. more than the U.K. cites Canada
U.S. – U.S.S.R.:	(a)	U.S. cites itself more than the U.S.S.R. cites itself
	(b)	U.S.S.R. cites the U.S. more than the U.S. cites the U.S.S.R.
	(c)	U.S. cites the U.K., Sweden and Switzerland more than the U.S.S.R. does

Table 6
Analysis of citation patterns for journals with 15 or more citations, corrected for publishing patterns (Fig. 2)

U.S. – Germany:	(a)	U.S. cites itself more than Germany cites itself
	(b)	Germany cites the U.S. more than the U.S. cites Germany
	(c)	U.K. cites the U.S. more than Germany
U.S. – U.S.S.R.:	(a)	U.S. cites the U.S.S.R. more than U.S.S.R. cites the U.S.
	(b)	U.K., France, Canada, Japan, Switzerland and Germany cite the U.S. more than they cite the U.S.S.R.
U.S. – Canada:	(a)	U.S. cites itself more than Canada cites itself
U.K. – Canada:	(a)	U.K. cites itself more than Canada cites itself
	(b)	U.K. cites Canada more than Canada cites the U.K.
	(c)	U.S. cites Canada more than the U.K. cites Canada
U.S.S.R. – Canada:	(a)	U.S.S.R. cites itself more than Canada cites itself
	(b)	U.K. and the U.S. cite Canada more than the U.S.S.R.
Canada – France:	(a)	France cites itself more than Canada cites itself
	(b)	U.S. cites Canada more than France cites Canada
U.S.S.R – Germany:	(a)	U.S.S.R. cites itself more than Germany cites itself
	(b)	U.S., the U.K., and Switzerland cite Germany more than they do the U.S.S.R.

Table 7
Analysis of citation patterns for recent citations, corrected for publishing patterns (Fig. 3)

U.S. – Germany:	(a)	U.S. cites itself more than Germany cites itself
	(b)	U.S. cites Germany more than Germany cites the U.S.
	(c)	Canada, Italy and Australia cite the U.S. more than does Germany
U.S. – U.S.S.R.:	(a)	U.K. and Canada cite the U.S. more than does the U.S.S.R.
U.K. – Canada:	(a)	U.K. cites Canada more than Canada cites the U.K.
U.S.S.R. – Canada:	(a)	U.S.S.R. cites itself more than Canada cites itself
	(b)	U.S. and the U.K. cite Canada more than does the U.S.S.R.
U.S.S.R. – Germany:	(a)	U.S.S.R. cites itself more than Germany cites itself
	(b)	U.S. cites Germany more than does the U.S.S.R.
U.S. – Canada:	(a)	U.S. cites itself more than Canada cites itself
	(b)	U.S. cites Canada more than Canada cites the U.S.

sed. We see again a high degree of self-citation in the U.S., even when publishing pattern or the degree of scientific currency is taken into account. As well, other countries tended, in 1969 at least, to circle the U.S. like planets around a sun in terms of their citation patterns. Studies by D. de SOLLA PRICE of Yale have indicated that this pattern is changing towards one of less concentration, and it will be of interest to repeat this study for a later time period.

We can view the relationships of Tables 5–7 as "building blocks" in devising scientific hierarchies among nations. The Computer Horizons group of Cherry Hill, N.J., U.S.A. has done pioneering work in hierarchical structures both among nations and areas of science.[42, 43] The present section may be viewed as a contribution to this field.

Conclusions

The input-output matrix may be adapted from its original use in economics to studies in science. We can use it to find both the relationships between pairs of countries and the interrelationships between many countries, using multi-dimensional scaling. We have found that the U.S. exerts a large influence on world science, both in terms of publishing and in citations to its papers. The relationships among other countries are less distinct, and depend on whether one is considering publications, all citations or recent citations.

*

We are indebted F. NARIN for data, and to D. de SOLLA PRICE and S. GURSEY for useful comments and access to data. The work was perfomed while H. I. was a Visiting Lecturer at the Department of History of Science and Medicine in Yale University, New Haven.

Appendix

Suppose there are k countries whose citation patterns we wish to compare. Let P_{ij} be true but unknown proportion of citations made by country i to country j. It is clear that $P_{i1} + P_{i2} + \ldots P_{ik} = 1$ for all $i = 1, 2, \ldots k$.

Let x_{ij} be the observed number of citations made by country i to country j in a sample of n_i citations. The vector $(x_{i1}, x_{i2} \ldots x_{ik})$ for fixed index i has a multinomial distribution with parameters $(p_{i1}, p_{i2} \ldots p_{ik}; n_i)$. Assume that the distribution of this vector can be approximated by a multivariate normal distribution and that data vectors for different countries are independently distributed. It is then possible to make use of a result, due to GOLD, to the effect that simultaneous confidence intervals can be constructed for all linear combinations $Q = \sum_i \sum_j c_{ij}\, p_{ij}$, with a prescribed level of confidence.

We assert that for all vectors $c = (c_{11}, c_{12} \ldots c_{kk})$ the sum Q is included in the interval

$$\sum_i \sum_j c_{ij}\, p_{ij} \pm [\chi^2_{k(k-1)}(\alpha)]^{1/2}\, s_c$$

with a confidence probability of $(1-\alpha)$, where

$$s_c^2 = \sum_{i=1}^{k} \frac{1}{n_i} \left[\sum_{j=1}^{k} c_{ij}^2\, p_{ij} - \left(\sum_{j=1}^{k} c_{ij}\, \hat{p}_{ij} \right)^2 \right]$$

$$\hat{p}_{ij} = x_{ij}/n_i$$

and $\chi^2_{k(k-1)}(\alpha)$ is the upper 100 $(1-\alpha)$ percentage point of a chi-square distribution with $k(k-1)$ degrees of freedom. We shall assume $\alpha = 0.1$.

To illustrate this result, suppose that we compare the self-citation pattern of countries 1 and 2. By choosing $c_{11} = 1$, $c_{22} = -1$ and $c_{ij} = 0$ for all other choices of i and j, the confidence interval for $p_{11} - p_{22}$ is

$$(\hat{p}_{11} - \hat{p}_{22}) \pm [\chi^2_{k(k-1)}(\alpha)]^{1/2}\, s_c$$

where

$$s_c^2 = \frac{\hat{p}_{11}(1-\hat{p}_{11})}{n_2} + \frac{\hat{p}_{22}(1-\hat{p}_{22})}{n_2}.$$

If this confidence interval excludes 0, we conclude p_{11} and p_{22} are statistically significant in their differences at the given level. Countries 1 and 2 then differ with respect to their self-citation proportions.

Suppose we wish to compare the proportion of citations that countries 1 and 2 make to one another. By selecting $c_{12} = 1$, $c_{21} = -1$ and $c_{ij} = 0$ for all other i, j, we deduce that a confidence interval for $p_{12} - p_{21}$ is

$$\hat{p}_{12} - \hat{p}_{21} \pm [\chi^2_{k(k-1)}(\alpha)]^{1/2}\, s_c$$

where

$$s_c^2 = \frac{\hat{p}_{12}(1-\hat{p}_{12})}{n_1} + \frac{\hat{p}_{21}(1-\hat{p}_{21})}{n_2}.$$

This procedure allows us to make any particular comparison of proportions. A difference is significant if the corresponding confidence interval does not include zero. The collection of all confidence intervals obtained has the confidence level $1-\alpha$.

While this approach is precise, it can be laborious. With 18 countries, we would have 153 possible pairs. With 18 comparisons per pair, we have to construct 2754 confidence intervals per matrix. We may argue that only those countries whose overall pattern is greatly different should be considered in detail. Multidimensional scaling,[37,38] provides a graphical means of detecting gross differences. Use of this technique requires the specification of a distance measure as well as a measure of similarity between vectors of the form $\hat{p}_i = \hat{p}_{il}, \ldots \hat{p}_{ik}$
where

$$\hat{p}_i = \frac{x_{il}}{n_i}$$

We define the distance measure as

$$d_{ij} = [(\hat{p}_{ij} - \hat{p}_{ji})^2 + (\hat{p}_{ii} - \hat{p}_{jj})^2 + \sum_{\substack{l \neq i \\ l \neq j}} (\hat{p}_{il} - \hat{p}_{jl})^2]^{1/2}$$

For the measure of similarity δ_{ij}, we select the test statistic which one would use in testing the hypothesis that the population proportions are the same:

$$\delta_{ij} = \left\{n_i \sum_{l=1}^{k} [(\hat{p}_{il} - q_{il})^2/q_{il}] + n_j \sum_{l=1}^{k} [(\hat{p}_{jl} - q_{jl})^2/q_{jl}]\right\}$$

where $q_{il} = q_{jl} = (x_{il} + x_{jl})/(n_i + n_j)$
for $l \neq i,\ l \neq j,\ l = 1 \ldots k$

$q_{ij} = q_{ji} = (x_{ij} + x_{ji})/(n_i + n_j)$

$q_{ii} = q_{jj} = (x_{ii} + x_{jj})/(n_i + n_j)$

The likelihood ratio test would reject the hypothesis for large values of δ. The multidimensional scaling program reorients the input-output matrix into a two-dimensional configuration of points so that vectors whose similarity measure is small are closer together than those for which it is not.

References

1. E.J. KNAPTON, *France: An Interpretive History,* Charles Scribner's Sons, New York, 1971, p. 253.
2. S. B. AHMED, Input-Output Analysis in Environmental Monitoring, *IEEE Transactions on Systems, Man and Cybernetics,* SMC3, 6 (1973) 537–8.
3. K. CHEN, Input-Output Economic Analysis of Environmental Impact, ibid, SMC 3, 6 (1973) 539–547.
4. I. GUTMANIS, Environmental Implications of Economic Growth in the U.S. 1970 to 2000 – Input-Output Analysis, ibid, 6 (1973) 568–574.
5. L. D. MAXIM, C. L. BRAZIE, Multistage Input-Output Model for Evaluation of Environmental Impact of Energy Systems, ibid, 6 (1973) 583–587.
6. W. H. MIERNYK, Regional Input-Output Pollution Abatement Modell, ibid 6 (1972) 575–577.
7. T. PAGE, Pollution Affecting Producers in an Input-Output Context, Ibid, 6 (1973) 555–561.
8. D. SHETER, Forecasting Industrial Pollution in Haifa Bay Area With an Input-Output Model, *Socio-Economic Planning Sciences,* 7 (1973) No.5, 397–406.
9. A. S. COHEN, A. P. HURTER, Input-Output Analysis of Costs of Air Pollution Control, *Management Science,* 21B (1974) No. 4, 453–461.
10. F. GIARRATANI, Air Pollution Abatement-Output and Relative Price Effects – A Regional Input-Output Simulation, *Environment and Planning,* 6 (1974) No. 3, 307–312.
11. M. CHATTERJEE, Dynamic Balanced Regional Input-Output Model of Pollution Control, ibid, 7 (1975) 21–34.
12. W. A. FLICK, Environmental Repercussions and Economic Structure – Input-Output Approach – Comment, *Review of Economics and Statistics,* 56 (1974) No. 1, 107–109.
13. W. LEONTIEF, Environmental Repercussions and Economic Structure – Input-Output Approach – Reply, ibid, 56 (1974) No. 1, 109–110.
14. F. GIARRATANI, Air Pollution Abatement and Man-power Requirements – A Regional Input–Output Analysis, *Annals of Regional Science,* 9 (1975) No. 2, 83–90.
15. W. H. MIERNYK, J. T. SEARS, *Air Pollution Abatement and Regional Economic Development – Input-Output Analysis,* Heath, Lexington, Mass., 1974.
16. M. B. HOCKING, Chemical Input-Output Analysis of Municipal Solid-Waste Incineration, *Journal of Environmental Systems,* 5 (1975) No. 3, 163–183.
17. J. J. DOUGLAS, Use of a Modified Input-Output Multiplier for Land Use Evaluation, *Australian Journal of Agricultural Economics,* 17 (1973) No. 1, 68–75.
18, M. L. PARKER, Input-Output View of Agriculture in Papua-New Guinea, ibid, 18 (1974) No. 1. 32–39.
19. D. MULKEY, J. C. HITE, Input Analysis of Energy Use by South Carolina Agriculture, *American Journal of Agricultural Economics,* 56 (1974) No. 5, 1203.
20. G. SCHLUTER, Combining Input-Output and Regression Analysis in Projection Models – Application to Agriculture, ibid, 56 (1974) No. 5, 1207.

21. C. ESTRUP, Energy Consumption Analysis by Application of National Input-Output Tables, *Industial Marketing Management,* 3 (1974) No. 4, 193–200.
22. J. B. PENN, B. McCARL, O. DOERING, L. BRINK, Structural Input-Output Modelling of Food and Fiber System Under Conditions of Fuel Scarcity, *American Journal of Agricultural Economics,* 56 (1974) No. 5, 1203.
23. M. A. SWIFT, Fuel Substitution in Input-Output Models – Non-linear Programming Approach, *Operations Research,* 23 (S2) B427 (1975).
24. M. O. FILANI, Air Traffic Forecasting – Input-Output Technique Approach, *Regional Studies,* 7 (1973) No. 3, 333–338.
25. S. L. GRAY, J. R. McKEAN, D. D. ROHDY, Estimating Impact of Higher Education from Input-Output Models – Case Study, *Rocky Mountain Social Science Journal,* 12 (1975) No. 1. 1–10.
26. B. C. LIU, Regional Hospital Needs Projection – Input-Output Approach, *Socio-Economic Planning Sciences,* 10 (1976) No. 1, 37–42.
27. A. A. KONIUS, Commensuration of Labor by Means of an Input-Output Table, *Matekon,* 9 (1973) No. 4, 31–42.
28. P. C. NYSTROM, Input-Output Processes of Federal Trade Commission, *Administrative Science Quarterly,* 20 (1975) No. 1, 104–113.
29. C. P. BOURNE, The World's Technical Journal Literature: An Estimate of Volume, Origin, Language, Field, Indexing and Abstracting, *American Documentation,* (1962) 159–168.
30. F. NARIN, *Evaluative Bibliometrics,* Computer Horizons Inc., Cherry Hill, N. J., 1976.
31. Published by the Institute for Scientific Information, Philadelphia, Pa.
32. H. CASON, M. LUBOTSKY, The Influence and Dependence of Psychological Journals on Each Other, *Psychological Bulletin,* 33 (1936) 95–103.
33. A. K. GUPTA, Characteristics of Documents cited by Indian Physicists: A Case Study, *DRTC Seminar* (India), No. 7, Paper JB, 1969.
34. F. NARIN, M. P. CARPENTER, National Publication and Citation Comparisons, *Journal of the American Society for Information Science,* 26 (1975) 80–93.
35. L. V. XHIGNESSE, C. E. OSGOOD, Bibliographic Citation Characteristics of the Psychological Journal Network in 1950 and in 1960, *American Psychologist,* 22 (1967) 778–791.
36. R. S. DANIEL, C. M. LOUTTIT, *Professional Problems in Psychology,* Prentice-Hall, New York, 1953.
37. J. B. KRUSKAL, Multidimensional Scaling by Optimizing Goodness of Fit to a Non-metric Hypothesis, *Psychometrika,* 29 (1964) No.1, 1–27.
38. J. B. KRUSKAL, Nonmetric Multidimensional Scaling: A Numerical Method, *Psychometrika,* 29 (1964) No. 2, 115–129.
39. H. INHABER, Where Scientists Publish, *Social Studies of Science,* 7 (1977) No. 3, 388–394.
40. Personal communication, F. NARIN, President, Computer Horizons Inc., Cherry Hill, N. J., Nov. 12, 1975.
41. H. INHABER, The Leading Edge of Science in Canada, *Research Policy,* 7 (1978) 88–98.
42. G. PINSKI, Influence and Interrelationship of Chemical Journals, *Journal of Chemical Information and Computer Sciences,* 17 (1977) 67–74.
43. J. D. FRAME, F. NARIN, M. P. CARPENTER, The Distribution of World Science, *Social Studies of Science,* 7 (1977) No. 4, 501–516.