To Contents Page

Chapter Six

A Science-Management Tool

Although it was developed primarily for bibliographic purposes, and in spite of its recognized utility as a search tool, the most important application of citation indexing may prove to be nonbibliographic. If the literature of science reflects the activities of science, a comprehensive, multidisciplinary citation index can provide an interesting view of these activities. This view can shed some useful light on both the structure of science and the process of scientific development. In this regard, the *SCI* data base is being used to do such things as evaluate the research role of individual journals, scientists, organizations, and communities; define the relationship between journals and between journals and fields of study; measure the impact of current research; provide early warnings of important, new interdisciplinary relationships; spot fields of study whose rate of progress suddenly begins accelerating; and define the sequence of developments that led to major scientific advances.

What the *SCI* data base brings to those kinds of problems is the ability to define two measures of scientific activity: the citation rates (how often cited) of authors, papers, and journals and the number of citation links between both given papers and given journals. Quantitative, objective, and fundamental, these measures are useful tools in managing science—not in the detailed sense of defining research objectives, routes, and timetables, but in the general sense of allocating resources and measuring progress.

QUALITATIVE MEASURE

The science-mangement applications of the *SCI* data base began with the simplest measure: citation rates. Underlying their use is the obvious need for some objective measure of the contributions made by individual researchers, papers, journals, programs, regions, and nations. Because they reflect the number of times individual scientists consider a given document important enough to cite it in their own work, citation rates appear to be such a measure. Certainly they provide the qualitative factor that is so glaringly absent from simple publication counts. By weighting individual publications on the basis of use by the scientific community, they add an important qualitative dimension to the publication record that is generally accepted as an analog, though an imperfect one, of research effort.

Admittedly, the nature of the quality that citation rates measure is elusive. It has been described variously as "significance," "impact," "utility," and "effectiveness," but no one has succeeded in defining it in more tangible terms. Nevertheless, two things are known about the quality that citation rates measure that make citation analysis a useful technique. One is that it is a positive quality; it generally reflects credit on the scientific work involved. The other is that it plays a significant role in the formation of peer opinions. The existence of these two characteristics are derived from a sizeable number of studies that show a strong, positive correlation between citation rates and peer judgments.

There are a number of theoretical objections raised about the use of citation rates as a measure of scientific quality (1). A person's rate could conceivably be inflated by self-citations. A paper might be cited frequently in refutation or as a negative example. There is no precise way of relating the citation rate of a co-authored paper to the contributions made by individual authors. A prestigious journal might draw more citations than a less prestigious one by providing more visibility. Primordial papers on methods that have been widely adopted tend to be cited with uncommon frequency, though it is debatable whether a methodological contribution is as important as a new theory, a conceptual insight, or an experimental finding. Then too, there is the problem of sloppy, and even biased, bibliographic practices (2). Not everyone cites all the obvious, classical antecedents or is conscientious about citing all the sources actually used. Not everyone conducts an exhaustive literature search or uses all the sources that should have been used. Not everyone limits references to only material that was actually read. And not everyone is objective about who is cited: some people cite a publication to make a friend look better, to flatter a superior, or to wrap themselves in the cloak of scholarship.

The validation studies done were designed to determine whether these factors negated, in fact as well as theory, citation rates as a general measure of scientific quality. The basic studies concentrated on papers and authors, which are closely intertwined. Any measure of the utility of a paper is also implicitly a measure of the work the author has done and is reporting. Since there is no other objective measure of scientific quality, the studies compared the judgments inferred from citation rates with the various forms of subjective peer judgments.

A study performed at ISI used the judgment of the Nobel Prize committees as a base line (3). The subjects of the study were the 1962 and 1963 winners of the Nobel Prize in physics, chemistry, and medicine. The rate at which their work had been cited was taken from the 1961 edition of *SCI* to eliminate any influence the award might have on their popularity as cited authors.

We found that the work of these authors was cited 30 times more frequently than the average for their fields. The average rate was 5.51 citations per author, as compared with 169 citations per Nobel Prize winner. Since Nobel Prize winners tend to publish more frequently than other scientists, we discounted the effect of frequency on the total rate by working out the average citation rate per paper for each author.

The Nobel Prize winners had an average citation rate per paper of 2.9, whereas their colleagues' rate was 1.57. The study showed, therefore, that quality judgments based on citation counts correlate very well with the judgments made by the Nobel Prize committees.

This study was repeated and extended in 1977, when we compiled the 1961–1975 citation rates of all Nobel Prize winners in science since 1950 (4). The list of Nobel laureates (see Figure 6.1) contained 162 names. The citation records of these scientists range from a high of 18,888 (L. D. Landau) to a low of 79 (J. H. D. Jensen), with a median rate of 1910. Only 6 of the laureates had citation counts under 200, and all of them did their award-winning research well before the advent of *SCI* in 1961. Thirty-eight received between 100 and 999 citations; 34, between 1000 and 1999; 21, between 2000 and 2999; 16, between 3000 and 3999; and 43 received over 4000 citations in the 15-year period. As a group they average 2877 citations. Taking the average of the authors listed in the 1970–1974 *SCI* cumulation for comparison purposes, we found that the average cited author could be expected to have accumulated less than 50 citations over the 15-year period.

In addition to determining whether honored scientists are also highly cited scientists, we looked at the corollary question too: Are highly cited scientists also honored scientists? Initially, we compiled a list of the 50 most-cited primary authors in 1967 (5). Six of them turned out to be Nobel Prize winners, and six more have been awarded Nobel Prizes since the study. In 1977 we extended the study to the 250 most-cited primary authors between 1961 and 1975 (4). Listed in Figure 6.2, 42 (17%) of them turned out to be Nobel laureates. Of the 250 most-cited primary authors, 151 (over 60%) have received the recognition of being elected to at least one national academy of science. Only 95 of them (38%) have won neither of these two honors. That doesn't mean, of course, that those 95 are unrecognized; if we had looked at the full range of awards, we most likely would have found that all of them had received recognition in one form or another.

K. E. Clark tested the accuracy of citation counts as a measure of quality in the field of psychology (6) by asking a panel of experts to list the people who they felt had made the most significant contribution to their specialties. He then measured the quality of the work done by the people listed by such criteria as citation counts, number of papers published, income, and number and quality of their students. The citation counts had the strongest correlation with the judgment of the panel.

Bayer and Folger used a sample of 467 biochemistry doctorates granted in 1967 and 1968 to determine, indirectly, how well citation counts correlated with peer judgments about the quality of educational institutions (7). The peer judgments were taken from a previous study in which a group of 152 biochemists were asked to rank the same departments that had granted the doctorates. Bayer and Folger then counted the citations received by each of the graduates and found that there was a strong correlation between the frequency of citation and the quality of the graduating institutions as ranked by professionals in the field.

Orr and Kassab compared citation rates against the peer judgments implied by the editorial rating of papers submitted for publication (8). The results were the same: there was a high correlation between citation counts and the judgments of the

PHYSICS

	Name	Country*	Total Citations 1961-1975
1950	Powell C	Britain	247
1951	Crockcroft JD	Britain	93
	Walton E	Ireland	112
1952	Bloch F	U.S.	2,188
	Purcell EM	U.S.	577
1953	Zernike F	Netherlands	467
1954	**Born M**	Germany	9,206
	Bothe W	Germany	201
1955	Kusch P	U.S.	459
	Lamb WE Jr.	U.S.	1,625
1956	**Bardeen J**	U.S.	4,788
	Brattain W	U.S.	303
	Shockley W	U.S.	3,571
1957	**Lee TD**	U.S.	4,879
	Yang CN	U.S.	1,728
1958	Cherenkov PA	U.S.S.R.	84
	Frank IM	U.S.S.R.	274
	Tamm IY	U.S.S.R.	1,144
1959	Chamberlain O	U.S.	236
	Segrè E	U.S.	493
1960	Glaser D	U.S.	343
1961	Hofstadter R	U.S.	1,686
	Mössbauer R	Germany	436
1962	**Landau LD**	U.S.S.R.	18,888
1963	Jensen JHD	Germany	79
	Mayer MG	U.S.	290
	Wigner EP	U.S.	4,948
1964	**Basov NG**	U.S.S.R.	4,320
1964	Prokhorov AM	U.S.S.R.	1,031
	Townes CH	U.S.	2,570
1965	**Feynman RP**	U.S.	6,031
	Schwinger JS	U.S.	4,855
	Tomonaga S	Japan	236
1966	Kastler A	France	570
1967	**Bethe HA**	U.S.	7,718
1968	Alvarez LW	U.S.	331
1969	**Gell-Mann M**	U.S.	9,669
1970	Alfvén HOG	Sweden	1,909
	Neel LEF	France	3,070
1971	Gabor D	Britain	1,749
1972	**Bardeen J**	U.S.	4,788
	Cooper LN	U.S.	323
	Schrieffer JR	U.S.	1,472
1973	Esaki L	Japan	747
	Giaever I	U.S.	695
	Josephson B	Britain	1,265
1974	Hewish A	Britain	766
	Ryle M	Britain	890
1975	Bohr AN	Denmark	3,517
	Mottelson BR	Denmark	1,362
	Rainwater J	U.S.	300
1976	Richter B	U.S.	205
	Ting SCC	U.S.	303
1977	**Anderson PW**	U.S.	6,787
	Mott NF	Britain	10,473
	Van Vleck JH	U.S.	5,449

CHEMISTRY

1950	Alder K	Germany	4,450
	Diels O	Germany	1,372
1951	McMillan EM	U.S.	97
	Seaborg G	U.S.	638
1952	Martin AJP	Britain	777
	Synge R	Britain	417
1953	Staudinger H	Germany	3,325
1954	**Pauling LC**	U.S.	15,662
1955	Du Vigneaud V	U.S.	1,470
1956	Hinshelwood C	Britain	476
	Semenov N	U.S.S.R.	1,257
1957	Todd A	Britain	275
1958	Sanger F	Britain	3,716
1959	Heyrovsky J	Czech	1,418
1960	Libby WF	U.S.	832
1961	Calvin M	U.S.	2,713
1962	Kendrew JC	Britain	1,654
	Perutz MF	Britain	4,263
1963	**Natta G**	Italy	5,735
	Ziegler K	Germany	3,258
1964	Hodgkin DMC	Britain	359
1965	**Woodward RB**	U.S.	7,069
1966	**Mulliken RS**	U.S.	10,508
1967	**Eigen M**	Germany	4,980
	Norrish RGW	Britain	980
	Porter G	Britain	3,202

Figure 6.1 Nobel prize winners since 1950 in physics, chemistry, and physiology or medicine. Total citations from 1961 to 1975 based on data from the *Science Citation Index*. Names in bold type also rank among the 250 most cited primary authors from 1961 to 1975.

CHEMISTRY (continued)

	Name	Country*	Total Citations 1961-1975
1968	Onsager L	U.S.	3,569
1969	**Barton DHR**	Britain	8,135
	Hassel O	Norway	1,113
1970	Leloir LF	Argentina	2,221
1971	Herzberg G	Canada	13,110
1972	Anfinsen CB	U.S.	2,286
	Moore S	U.S.	8,167
	Stein WH	U.S.	1,274
1973	**Fischer E**	Germany	4,788
	Wilkinson G	Britain	967
1974	**Flory PJ**	U.S.	10,247
1975	Cornforth JW	Australia	2,378
	Prelog V	Switzerland	2,229
1976	Lipscomb WN	U.S.	1,443
1977	**Prigogine I**	Belgium	4,681

PHYSIOLOGY OR MEDICINE

1950	Hench PS	U.S.	316
	Kendall EC	U.S.	179
	Reichstein T	Switzerland	1,178
1951	Theiler M	South Africa	206
1952	Waksman SA	U.S.	2,291
1953	Lipmann FA	U.S.	2,038
	Krebs HA	Britain	7,657
1954	Enders JF	U.S.	1,193
	Robbins FC	U.S.	584
	Weller TH	U.S.	1,972
1955	Theorell AHT	Sweden	3,150
1956	Cournand AF	U.S.	1,263
	Forssmann W	Germany	637
	Richards D	U.S.	668
1957	Bovet D	Italy	1,219
1958	Beadle GW	U.S.	948
	Lederberg J	U.S.	3,138
	Tatum EL	U.S.	285
1959	**Kornberg A**	U.S.	4,548
	Ochoa S	U.S.	2,425
1960	**Burnet FM**	Australia	5,553
	Medawar PB	Britain	2,600
1961	von Békésy G	U.S.	1,960
1962	Crick FHC	Britain	2,524
	Watson JD	U.S.	2,437
	Wilkins MHF	Britain	745
1963	**Eccles JC**	Australia	10,104
	Hodgkin AL	Britain	7,500
	Huxley AF	Britain	2,115
1964	Bloch K	U.S.	1,456
	Lynen F	Germany	3,020
1965	**Jacob F**	France	7,101
	Lwoff A	France	2,111
1965	**Monod J**	France	4,791
1966	Huggins CB	U.S.	3,808
	Rous FP	U.S.	1,396
1967	**Granit RA**	Sweden	4,629
	Hartline HK	U.S.	1,183
	Wald G	U.S.	3,002
1968	Holley RW	U.S.	2,296
	Khorana HG	U.S.	1,651
	Nirenberg MW	U.S.	1,916
1969	Delbruck M	U.S.	498
	Hershey AD	U.S.	2,039
	Luria SE	U.S.	1,876
1970	**Axelrod J**	U.S.	6,973
	Katz B	Britain	4,690
	von Euler U	Sweden	8,728
1971	**Sutherland EW**	U.S.	5,150
1972	Edelman GM	U.S.	3,414
	Porter RR	Britain	2,528
1973	von Frisch K	Germany	955
	Lorenz KZ	Germany	1,560
	Tinbergen N	Netherlands	1,205
1974	**DeDuve C**	Belgium	8,445
	Claude A	U.S.	493
	Palade GE	U.S.	5,969
1975	Baltimore D	U.S.	2,543
	Dulbecco R	U.S.	4,005
	Temin HM	U.S.	3,168
1976	Blumberg BS	U.S.	3,555
	Gajdusek DC	U.S.	1,318
1977	Guillemin R	U.S.	2,395
	Schally A	U.S.	2,985
	Yalow R	U.S.	3,658

*** Citizenship of recipient at time of award.**

Figure 6.1 (continued)

Name	Total Citations 1961-1975	National Academy	Name	Total Citations 1961-1975	National Academy
Abragam A	6,769	France	Brodie BB	7,493	U.S.
Abramowitz M	5,108		Brown HC	16,623	U.S.
Abrikosov AA	5,429	U.S.S.R.	Brown JB	4,074	
Albert A	8,664		Buckingham AD	4,332	U.K.
Allinger NL	4,140		Budzikiewicz H	5,089	
Allison AC	6,105		Bunnett JF	4,370	
Anden NE	5,147		Burn JH	5,650	U.K.
Anderson PW (77P)	6,787	U.S.	**Burnet FM** (60M)	5,553	U.K., U.S.
Andrews P	4,485		Burton K	6,913	U.K.
Arnon DI	4,323	U.S.	Busing WR	5,066	
Axelrod J (70M)	6,973	U.S.	Carlson LA	4,282	
Baker BR	5,395		Carlsson A	7,697	
Bardeen J (56P)	4,788	U.S., U.K.	Cattell RB	4,190	
(72P)			Chance B	16,306	U.S.
Barrer RM	5,230	U.K.	Chandrasekhar S	8,179	U.S., U.K.
Bartlett PD	5,180	U.S.	Chapman S	5,235	U.K., U.S.
Barton DHR (69C)	7,763	U.K., U.S.	Chatt J	6,692	U.K.
Basolo F	4,083		Clementi E	5,684	
Basov NG (64 P)	4,320	U.S.S.R.	Cohen MH	4,808	
Bates DR	6,925	U.K.	Conney AH	5,151	
Bell RP	4,400	U.K., U.S.	Cope AC	5,269	
Bellamy LJ	10,736		Corey EJ	9,901	U.S.
Bellman RE	5,678		Cotton FA	12,901	U.S.
Bender ML	4,924	U.S.	Coulson CA	6,569	U.K.
Benson SW	5,319		Courant R	4,154	
Bergstrom S	4,473	Sweden, U.S.	Cram DJ	6,148	U.S.
Berson SA	4,486		Cromer DT	5,418	
Bethe HA (67P)	7,718	U.S., U.K.	Cruickshank DWJ	4,512	
Beutler E	5,636	U.S.	Cuatrecasas P	4,484	
Billingham RE	6 269	U.K.	Curtis DR	4,794	
Birch AJ	4,339	U.K.	Dacie JV	4,323	U.K.
Bjorken JD	4,264	U.S.	Dalgarno A	5,365	U.K.
Bloembergen N	5,234	U.S.	Davis BJ	7,074	
Born M (54 P)	9,206	U.S.	Dawson RMC	4,125	
Bourbaki N	4,860		**DeDuve C** (74 M)	8,445	U.S., Belgium
Boyer PD	6,906	U.S.	DeRobertis E	4,801	
Brachet J	5,956	U.S., U.K.	Dewar MJS	9,800	U.K.
		France	Dische Z	7,874	U.S.
Braunwald E	4,980	U.S.	Dixon M	6,331	U.K.
Bray GA	8,012		Djerassi C	8,520	U.S.
Bridgman PW (46P)	5,053	U.S., U.K.	Doering WVE	4,253	U.S.

CONTINUED

Figure 6.2 Incidence of Nobel Prizes and memberships in national academies of science among the 250 most cited primary authors from 1961 to 1975. Citation rates are based on data from the *Science Citation Index*. Nobel laureates appear in bold type, followed by year and category of prize: P = physics, C = chemistry, M = physiology or medicine. Membership in national academies of science include correspondents, fellows, foreign members, and foreign associates.

Name	**Total Citations 1961-1975**	**National Academy**
Dole VP	5,902	U.S.
Duncan DB	4,153	
Eagle H	6,498	U.S.
Eccles J C (63M)	10,104	U.K., U.S.
Eigen M (67C)	4,980	U.K., U.S.
Eliel EL	8,615	U.S.
Erdelyi A	5,978	U.K.
Eysenck HJ	5,241	
Fahey JL	4,724	
Falck B	4,275	
Farquhar MG	4,525	
Fawcett DW	6,236	U.S.
Feigl F	4,074	
Feldberg W	4,762	U.K.
Feynman RP (65P)	6,031	U.S., U.K
Fieser LF	9,392	U.S.
Fischer EO (73C)	4,788	
Fisher ME	4,289	U.K.
Fisher RA	8,336	U.K.
Fiske CH	8,249	
Flory PJ (74 C)	10,247	U.S.
Folch J	9,693	
Fraenkel-Conrat H	4,376	U.S.
Fredrickson DS	6,897	U.S.
Freud S	8,490	U.K.
Friedel J	4,325	France
Gell-Mann M (69 P)	9,669	U.S.
Gilman H	7,849	U.S., U.K.
Ginzburg VL	6,834	U.S.S.R.
Glasstone S	5,080	
Gomori G	7,136	
Good RA	4,607	U.S.
Goodman LS	5,627	U.S.
Goodwin TW	4,727	U.K.
Gornall AG	5,921	Canada
Grabar P	4,717	
Granit RA (67 M)	4,629	U.K., U.S., Sweden
Green DE	4,708	U.S.
Gutowsky HS	4,286	U.S.
Hansen M	5,262	U.S.
Harned HS	4,960	U.S.
Herbert V	4,106	
Herzberg G (71 C)	13,110	U.S., U.K., Canada
Hirs CHW	4,578	
Hirschfelder JO	7,033	U.S.
Hodgkin AL (63 M)	7,500	U.K., U.S.
Horner L	4,469	
House HO	4,393	
Hubel DH	4,640	U.S.
Huisgen R	9,309	F.R.G., G.D.R.
Huxley HE	4,073	U.K.
Ingold CK	4,198	
Jackman LM	4,927	
Jacob F (65 M)	7,101	U.K. U.S., France
Jaffé HH	5,106	
Johnson HL	4,117	U.S.
Jorgensen CK	6,049	Denmark
Kabat EA	7,529	U.S.
Karnovsky MJ	5,616	
Karplus M	5,770	U.S.
Kato T	4,138	
Katritzky AR	4,704	
Katz B (70 M)	4,690	U.K., U.S.
Keilin D	4,121	
Kety SS	4,594	U.S.
King RB	5,109	
Kirkwood JG	4,084	U.S.
Kittel C	5,591	U.S.
Klein G	4,430	U.S.
Klotz IM	4,151	U.S.
Kolthoff IM	9,697	U.S.
Kornberg A (59 M)	4,548	U.S., U.K.
Krebs HA (53 M)	7,.657	U.K., U.S.
Kubo R	4,232	U.S.
Kuhn R (38C)	7,488	
Landau LD (62 P)	18,888	U.S.S.R.
Lee T D (57 P)	4,879	U.S.
Lehninger AL	5,507	U.S.
Lemieux RU	4,619	Canada, U.K.
Levine S	4,035	
Lineweaver H	5,202	
Löwdin PO	5,060	Sweden, Norway
Lowry OH	58,304	U.S.
Luft JH	8,926	
Marmur J	6,475	
McConnell HM	5,490	U.S.

Figure 6.2 (continued)

referees on a total of 5000 documents that had been submitted to two biomedical journals over a period of five years.

Still another test of the correlation between peer judgments of quality and citation rates was performed by Virgo (9). Her study had nine subject experts, from the

Name	Total Citations 1961-1975	National Academy
McKusick VA	4,181	
Miller JFA	6,371	U.K.
Millonig G	4,106	
Mitchell P	4,086	U.K.
Monod J (65 M)	4,791	U.S.
Moore S (72 C)	8,167	U.S.
Morse PM	5,089	U.S.
Mott NF (77 P)	10,473	U.K., U.S.
Muller A	4,500	
Müller E	4,664	U.S.
Mulliken RS (66 C)	10,508	U.S., U.K.
Nakamoto K	5,132	
Natta G (63 C)	5,735	Italy, France, U.S.S.R.
Nesmeyanov AN	6,783	U.S.S.R., U.K.
Newman MS	4,730	U.S.
Novikoff AB	7,662	U.S.
Olah GA	8,311	U.S.
Ouchterlony O	5,986	
Palade GE (74 M)	5,969	
Pauling L (54 C) (62 Peace)	15,662	U.S., France, U.K., U.S.S.R.
Pearse AGE	10,522	
Perutz MF (62 C)	4,263	U.K., U.S., France
Pople JA	15,135	U.K.
Prigogine I (77C)	4,681	U.S.
Racker E	4,567	U.S.
Reed LJ	4,290	U.S.
Reynolds ES	10,115	
Roberts JD	4,501	U.S.
Robinson RA	5,543	
Rose ME	4,127	
Rossini FD	4,105	U.S.
Russell GA	5,933	
Sabatini DD	6,205	
Scatchard G	4,191	
Scheidegger JJ	4,159	
Schneider WC	7,029	
Schwarzenbach G	4,618	
Schwinger J (65 P)	4,855	U.S.
Seeger A	4,757	
Seitz F	5,396	U.S.
Selye H	8,928	Canada
Seyferth D	4,462	
Sillen LG	4,375	
Skou JC	4,127	
Slater JC	7,587	U.S.
Smith HW	6,946	
Smithies O	6,192	U.S.
Snedecor GW	14,762	
Somogyi M	4,465	
Spackman DH	6,889	
Spitzer L	4,238	U.S.
Stahl E	6,252	
Steel RGD	5,100	
Streitwieser A	7,511	U.S.
Sutherland EW (71 M)	5,150	
Taft RW	5,083	
Tanford C	5,934	U.S.
Udenfriend S	5,039	U.S.
Umbriet WW	5,229	
Van Slyke DD	4,282	
Van Vleck JH (77 P)	5,449	U.S., U.K., France
von Euler US (70M)	8,728	U.S., U.K.
Walling C	5,590	U.S.
Warburg O (31 M)	7,463	U.K.
Warren L	4,303	
Watson ML	4,176	
Weber G	8,319	U.S.
Weber K	5,823	
Weinberg S	6,306	U.S.
Weiss P	4,048	U.S.
Wiberg KB	5,461	U.S.
Wieland T	4,423	
Wigglesworth VB	4,489	U.K., U.S.
Wigner EP (63P)	4,948	U.S., U.K.
Wilson EB	5,139	U.S.
Winer BJ	5,145	
Winstein S	7,884	
Wittig G	6,079	France
Woodward RB (65 C)	7,069	U.S., U.K.
Zachariasen WH	4,050	U.S.
Zeldovich YB	4,794	U.S.S.R.
Ziman JM	4,499	U.K.
Zimmerman HE	4,217	

Figure 6.2 (continued)

fields of surgery and radiology, select papers relevant to their research and then rate five pairs of them in two different ways. The individual papers in each pair were ranked by relative importance, and all the papers were rated on a quality scale of 1 to 5. The citation frequency of each paper turned out to correlate with the relative ranking of each pair of papers at least as well as the ratings by a second set of subject experts. In addition, when Virgo attempted to determine which of 10 objective and seven subjective variables associated with the papers correlated with the scale rating of 1 to 5, she found that a combination of two citation measures was the only one that did. In fact, the correlation was even stronger than the one between citation frequency and the relative ranking of the pairs of papers. One of the measures was the citation rate achieved by the paper. The other was the average citation rate per item published achieved by the publishing journal.

A study conducted by ISI for the Air Force Office of Scientific Research (AFOSR) proved the same point about citation counts, while demonstrating one of the applications of the measure. The study consisted of counting the citations received during 1965–1966 by papers that were published in 1964 on research sponsored by the AFOSR. The purpose of the study was to see how well the AFOSR was doing in selecting projects to support; that was the application side of the coin. The other side was that the AFOSR had a rather rigorous process for screening the proposals it received for research support. They used outside referees, panels from inhouse laboratories, and other methods to select the best of the proposals. In other words, they had a rather elaborate system of peer judgments for measuring the quality of the scientists and proposals they considered supporting.

In comparing the citation counts of the papers published from AFOSR-supported research against the citation counts of a random sample of papers taken from the same journals, the study showed that the AFOSR-supported papers that were cited at least once drew an average of 2.10 citations versus 1.63 citations for the control papers that had been cited at least once. From the Air Force viewpoint, this finding confirmed the effectiveness of their selection process. From the viewpoint of those interested in an objective measure of scientific performance, however, it also confirmed, once again, that there is a high correlation between citation counts and peer judgments on the subject of scientific quality.

What all the studies show, therefore, is that of all the variables that can influence citation rates, the scientific quality of the work published is the dominant one. Sloppy, biased bibliographic practice is a random variable that tends to get canceled out. The same thing cannot be said for the variables of exposure, prestige, coauthors, and nature of the references. They are not random; they do not get canceled out. They must be considered in any citation analysis of a person or a paper, and they negate any quality judgments that might be made on the basis of small differences in citation rates. But the evidence shows that these variables are not strong enough in their influence to explain large differences in citation counts. Apparently, only differences in quality and impact account for that.

CITATION INSIGHTS

The ability of citation counts to provide a rough, but objective and useful, relative measure of scientific quality promises to have some profound implications. Following in the steps of the Air Force, other government agencies are using citation analysis to improve their ability to define what is going on in scientific fields of interest. A study conducted by ISI for the National Science Foundation on the characteristics of frequently cited papers in chemistry is typical.

Some of the main findings of the study were:

1. Seventy percent of the heavily cited (10 times or more in the year studied) items were published during the preceding 10 years.
2. The items most heavily cited, particularly by applied chemists, tended to be books that were published early in the 10-year time frame.
3. Theoretical papers dominated the list of the 50 most frequently cited items. Experimental methodology was the next most frequently cited type of subject matter.
4. The central specialty of chemistry was molecular orbital theory.
5. A high percentage of highly cited chemists were receiving NSF support, and the amount of support NSF was providing to the most highly cited ones was substantially higher than the average NSF award.

This view of the inner workings of the science of chemistry was enlightening enough for the National Science Foundation to extend its investigation to include the engineering sciences, and to take a closer look at the cross-disciplinary papers in chemistry.

Similar citation studies, looking at similar characteristics, have been conducted for other government agencies—including the National Institute of Mental Health, the National Cancer Institute, and the Consiglio Nationale delle Ricerche (CNR) in Italy. The National Institute of Mental Health was concerned with measuring the output of its research grants. The National Cancer Institute was looking for statistical data that would help it evaluate proposals for the support of cancer centers. CNR was looking for information on the life sciences that would have a bearing on a variety of science-policy decisions.

Citation rates of individual papers, or groups of papers that define given fields, are also being used to identify research areas marked by sudden spurts of activity. Price has used the *SCI* to develop an average citation-rate curve that can be used as a base line for spotting groups of papers whose rate is higher, increasing faster, or more enduring (10). A study of the literature on pulsars (11) suggests that these characteristics typify an emerging field.

At the opposite end of the spectrum from Price's macro views of science, citation counts are being used by others to provide a micro view of individual scientists. At least one major university has reversed a decision to refuse tenure after a citation analysis was done of the applicant's work. In another, much more public case (12), citation analysis was used to support a formal legal challenge of a tenure decision. The challenge came from a female biochemist, who was denied tenure at the same time it was granted to two male colleagues. She claims they are no better qualifed than she. The claim has been quantified by Robert E. Davies, a biochemist at the University of Pennsylvania, who was asked to testify as an expert witness on tenure procedures. Davies and two operations-research specialists, Nancy L. Geller and

John S. De Cani, have developed a way of estimating the lifetime citation rates of a given paper (13). The technique, Davies claims, is a careful one, which compensates for such disturbing variables as self-citation, derogatory citations, multiple authorships, prestige of the publishing journal, and tendency for papers on widely useful methods to be heavily cited. According to Davies and his colleagues, citation measures show that the research work of the biochemist who was denied tenure is superior to the two faculty members who received it and, in fact, is on a par with the full professors in the department. The evidence, however, did not prevent the court from rejecting the claim.

STRUCTURAL RELATIONSHIPS

Studies of the citation links between papers are providing still different views of science. In his work on the sociology of science, Price has shown that the distribution of references by the age of cited papers provides a way of distinguishing between hard science, soft science, and the humanities—each of which is built on a different social system and progresses in a different manner and at a different rate (14). He did this by developing an immediacy index, which describes the percent of total references that cite literature published in the last five years. In analyzing the material published by journals in a number of fields, he found that their immediacy-index rating agreed with intuitive judgments about what is hard science, soft science, and the humanities. Journals of physics and biochemistry have an immediacy index of 60 to 70%. Journals in the field of radiology show a 54 to 58% index value. The *American Sociological Review* has an index of 46.5%, and journals dealing with the study of literature as an art form are all less than 10%.

Earlier (10), Price had shown that the literature of any given field is made up of two segments: the archival literature and the recent literature that describes the research front of the field. His work on the age of references led him to conclude that the frequency with which authors cite the research-front literature is a measure of the "hardness" of the field.

Another way in which citation links between papers can help shed some light on the sociology of science is by providing a graphic, detailed picture of the history of major scientific developments. This application is not as far advanced as some of the other citation techniques used in sociological studies, but its potential is at least as great.

Working under a contract from the Air Force Office of Scientific Research, ISI has already used the *SCI* data base to construct a network diagram (see Figure 6.3) that defines the particular sequence of research events that culminated in the discovery of the DNA code (15). Definition is in terms of the key research events, their relative importance, chronological sequence, and relationships to each other. The key events were taken from the Asimov book, *The Genetic Code* (16). Each was represented by one or more of the published papers in which the research of the event was originally described. The relative importance of the events was measured by the number of times each paper was cited. The relationships between events were defined by the citation links between the papers representing the events.

To test the accuracy of the network, we constructed another diagram of the same development that was based entirely on the Asimov account. In this network, the relationships shown are those described by Asimov.

The citation network confirmed 65% of the relationships described by Asimov. And when the events in the citation network were weighted on the basis of the number and type of citation links, the one that scored the highest was the same one that Asimov judged to be the single most important contribution.

The citation analysis did more than just duplicate most of the account that Asimov had put together from a remarkable memory. It also added some insights into what happened by identifying 31 relationships and one event that Asimov did not mention. The event was identified not by the citation network, but by the citation index from which the network was developed. The index showed every paper cited by the papers representing the Asimov events. This view of the development process identified 26 authors who were cited by "event" authors, but who were not mentioned by Asimov. Four of those authors were cited for work that played an important role in the development and verification of the DNA-code theory. The work of at least one of the four seems to have been sufficiently critical to be included in the Asimov account.

The study made four significant points about the use of citation analysis for historical research. First, the relationships that a citation analysis shows among the components of a given body of work correspond very well to the relationships perceived by a scientist of Asimov's rank. Second, a citation analysis can identify significant relationships and events that even a remarkable memory might forget, or that traditional techniques of historical research might miss. Third, a graphic presentation of the sequence of events is superior to a narrative presentation for the purposes of historical and sociological analyses. And fourth, the manual construction of network diagrams, named "historiographs," was much too laborious for them to ever become widely used.

The last point led to additional research, which is still continuing, into the feasibility of computer-generated historiographs. Such a development is, at the very least, technologically feasible. Given the continuing rapid development of computer technology, economic feasibility looks promising, which makes it likely that some time in the intermediate future science historians and sociologists will be able to sit at a computer terminal and generate historiographs from a citation-index data base as easily as they now perform ordinary literature searches.

Citation links show just as useful a picture of the present as of the past. A research program at ISI is using citation links to graphically depict the high activity areas of science (17). Several types of citation counts also are involved in this process. Straight citation counts are used to identify the highly cited items in a given year. Co-citation counts, the number of times a pair of papers has been cited by individual source papers that year, are used to organize the papers into clusters and show the relationships between clusters. A cluster consists of all the papers linked by co-cita-

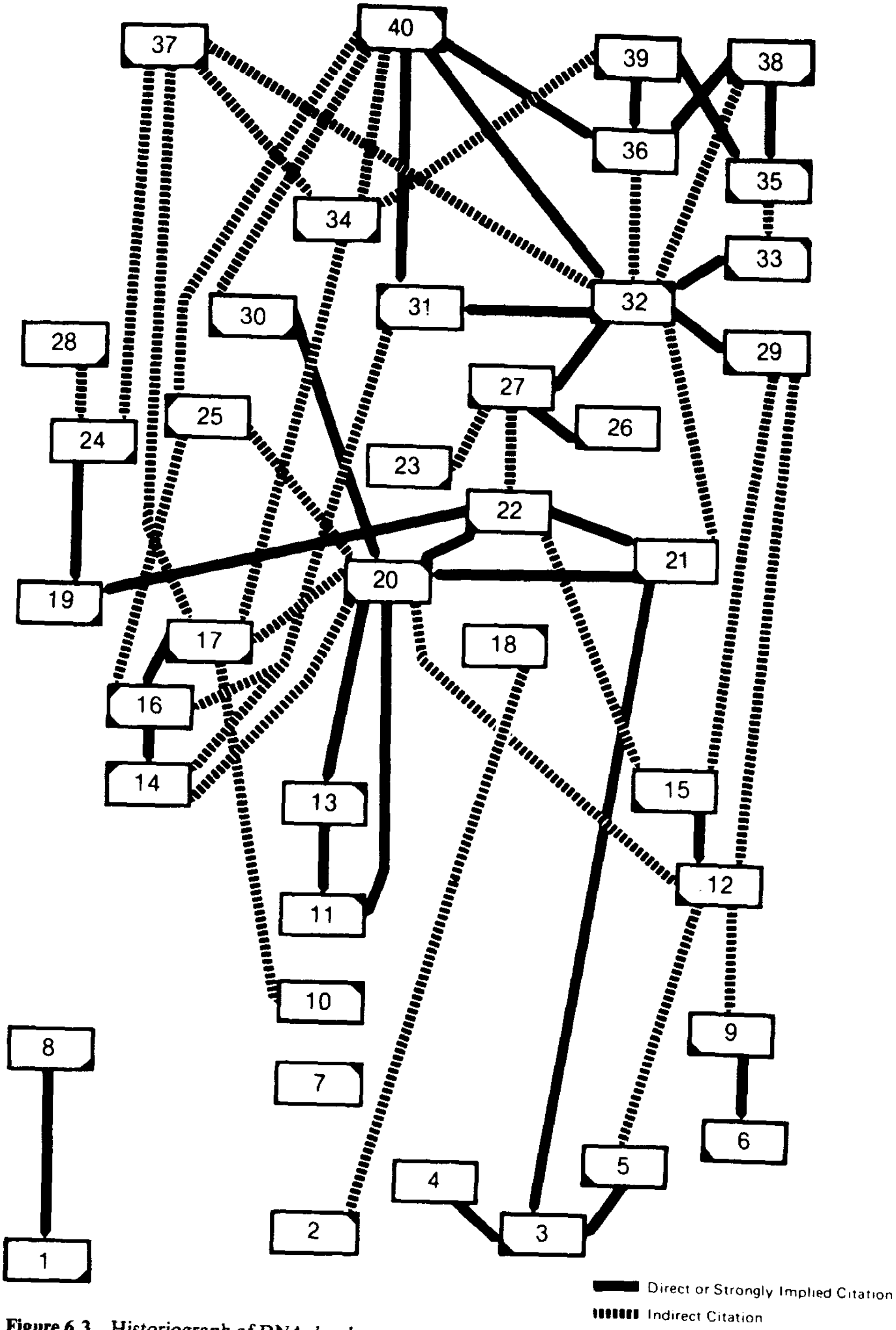

Figure 6.3 Historiograph of DNA development.

KEY

1. Braconnot 1820
2. Mendel 1865
3. Miescher 1871
4. Flemming 1879
5. Kossel 1886
6. Fischer and Piloty 1891
7. DeVries 1900
8. Fischer 1907
9. Levene and Jacobs 1909
10. Muller 1926
11. Griffith 1928
12. Levene with Mori and London 1929
13. Alloway 1932
14. Stanley 1935
15. Levene and Tipson 1935
16. Bawden and Pirie 1936-1937
17. Caspersson and Schultz 1938-1939
18. Beadle and Tatum 1941
19. Martin and Synge 1943-1944
20. Avery, MacLeod, and McCarty 1944
21. Chargaff 1947
22. Chargaff 1950
23. Pauling and Corey 1950-1951
24. Sanger 1951-1953
25. Hershey and Chase 1952
26. Wilkins 1953
27. Watson and Crick 1953
28. DuVigneaud 1953
29. Todd 1955
30. Palade 1954-1956
31. Fraenkel-Conrat 1955-1957
32. Ochoa 1955-1956
33. Kornberg 1956-1957
34. Hoagland 1957-1958
35. Jacob and Monod 1960-1961
36. Hurwitz 1960
37. Dintzis 1961
38. Novelli 1961-1962
39. Allfrey and Mirsky 1962
40. Nirenberg and Matthaei 1961-1962

tions at a frequency level equal to or greater than a given threshold. In other words, every paper in a cluster has been co-cited with at least one other paper in the cluster *n* times (threshold level) or more. Co-citation links below the threshold level are used to show the relationships between clusters.

When the titles of the papers in each cluster are analyzed, they are found to have certain words and concepts in common that suggest names descriptive of the type of research being reported. These cluster names seem to describe coherent scientific specialties. Authors of some of the cluster papers, with whom the names and contents of the clusters have been checked, confirmed that the names are, in fact, descriptive of their specialty and that the papers in the clusters represent the core literature of the specialty.

Except for the analysis of the paper titles and the naming of the clusters, this entire process is automatic. In other words, what we have developed is a computer model capable of mapping the structure of science in terms of its most active specialties. By changing the threshold levels of the citation and co-citation counts that qualify papers for inclusion in the model and its clusters, we can change the resolution of the map. The lower the threshold, the broader the view (see Figure 6.4); the higher the threshold, the narrower and sharper the view (see Figure 6.5).

The specialty viewpoint seems to be very useful. For one thing, it is detailed enough to be sensitive to the subtle changes that take place in scientific research from year to year. Maps of the biomedical group of clusters derived from the literature of 1972 and 1973 (Figure 6.6) showed significant changes in the relative importance of several specialties, shifts in the relationships between specialties, and the emergence of an important, new specialty.

The detail level and responsiveness of specialties seem well suited to a system for classifying scientific literature hierarchically by subject. The effectiveness of an indexing/retrieval system built on an hierarchy of subject classifications is a function

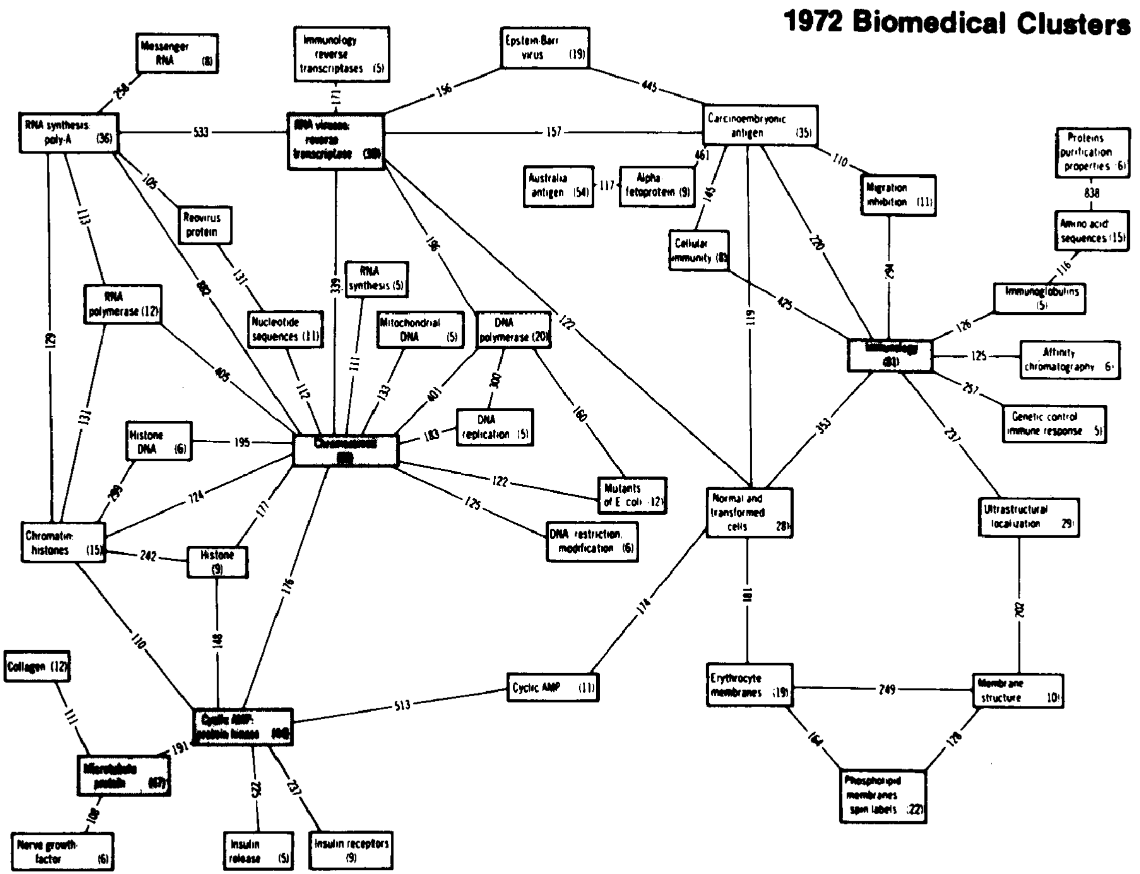

Figure 6.4 1972 Biomedical clusters. Each box represents a cluster of highly co-cited documents, which identify a particular specialty. The number in parentheses in each box indicates the number of co-cited documents in the cluster. The numbers on the lines connecting the boxes indicate the frequency with which documents in both clusters were co-cited.

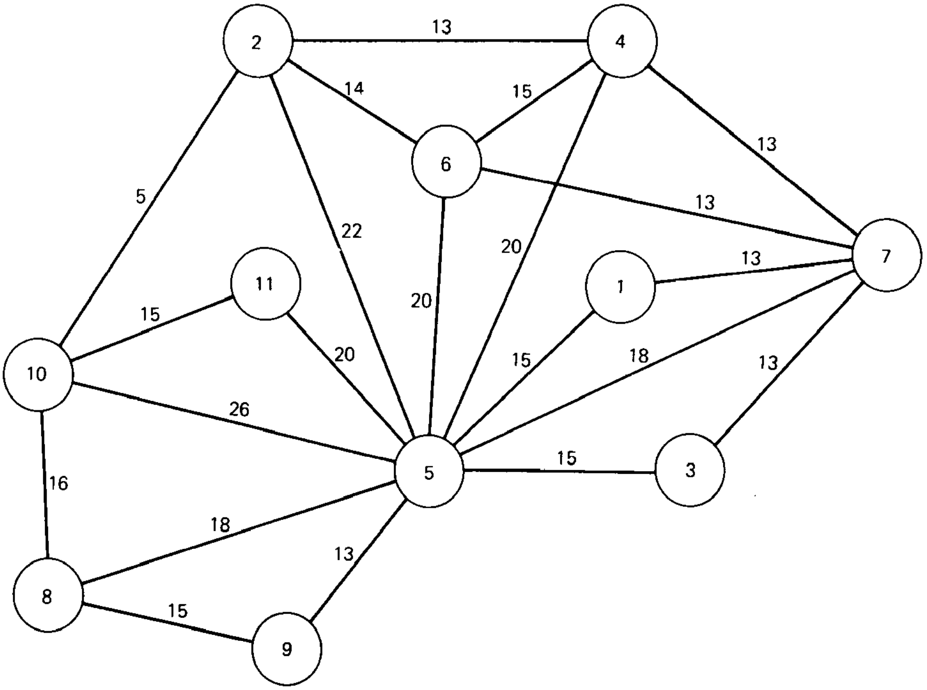

Figure 6.5 FSH and LH releasing hormone cluster in 1972. Each node represents a highly cited document. The numbers on the lines connecting pairs of nodes indicate the number of times the pairs of documents were co-cited.

KEY TO AUTHORS AND PUBLICATION YEARS OF NODAL PAPERS:

1. Amoss 1971
2. Baba 1971
3. Burbus 1971
4. Geiger 1971
5. Matsuo 1971
6. Matsuo 1971
7. Monahan 1971
8. Niswender 1968
9. Ramirez 1963
10. Schally 1971
11. Schally 1971

---

of how close the subject headings and their hierarchical relationships match reality. In the case of science, reality consists of the basic units of research and the relationships between them. Constructing such an hierarchy of descriptive terms is one of the primary difficulties in developing a useful classification system. Keeping the heirarchy current in the face of constant change is the other one.

The computer model of scientific specialties derived from the *SCI* data base may offer a way around these difficulties. The specialties defined by the model seem to be the basic units of research in the scientific process, and the relationships shown between them seem to correspond to the logical structure of the process. Equally as important, the automated nature of the model makes it practical to update the classification scheme yearly to keep pace with the dynamics of the process.

Theoretically, it should be possible to build an automatic classification system from the *SCI* model of scientific specialties. Such a system would automatically classify papers by their references, according to the cluster in which the reference

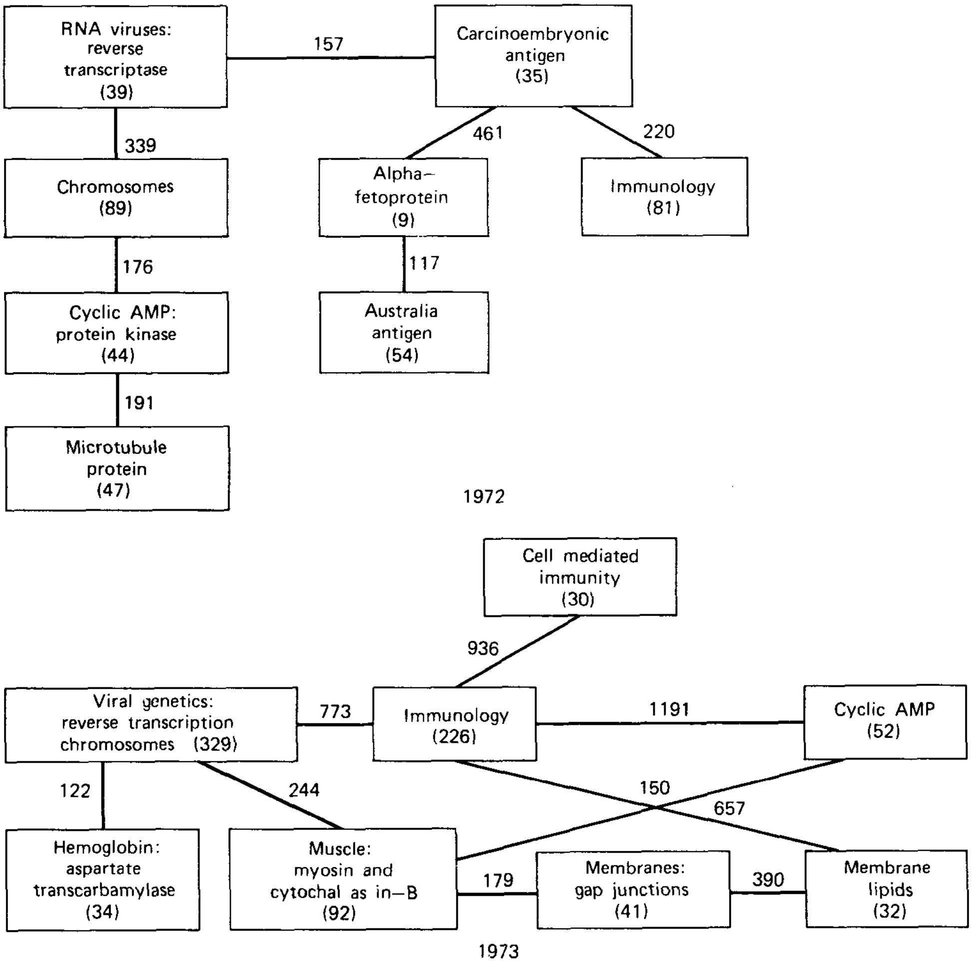

Figure 6.6 Major biomedical clusters in 1972 and 1973. Each box represents a cluster of highly co-cited documents on the subject of the specialty shown. The number in each box indicates the number of documents in the cluster. The numbers on the lines connecting boxes indicate the frequency with which documents in both clusters have been cited together.

citations are found. Research on such a system is being conducted at ISI.

There is one more area of science management in which citation counts and links are useful. The *SCI* data base shows measures not only for authors and papers, but also for journals. This information is published as a part of the *SCI* under the name *Journal Citation Reports.*

Journal Citation Reports® provides the following data on the source journals covered by *SCI:*

1. How often each journal is cited.
2. How many items it publishes.
3. How often (on an average) each item is cited, which is called "impact factor."
4. How often (on an average) each item is cited during the year of its publication, which is called "immediacy index."

5. The source journals responsible for the references to each journal, the number of references received from each, and how they were distributed by the publication years of cited issues.
6. The number of references each journal published, to what journals, and how the references were distributed by the publication years of cited issues.

As with all other citation measures, the ones given for journals are not absolute. Citation counts measure only one aspect of journal performance: that of disseminating research findings that are useful to scientists. They say nothing about a journal's performance in disseminating general news about a given area of scientific activity. And even at that, the citation counts can be influenced by such factors as the reputation of authors published, the controversiality of the subject matter, the journal's circulation, its reprint policies, and the coverage by indexing and abstracting services.

Nevertheless, as with authors and papers, a large difference in the citation counts of two journals indicates a significant difference in the quality of the research results they publish. Librarians concerned with the cost effectiveness of their journal collections, researchers and teachers who have to compile reading lists for given subject areas, journal editors looking for a way of measuring their performance against competition, and scientists doing research on one aspect or another of the scientific process, all find the journal citation counts useful.

For librarians and people doing general science studies, the citation links between journals are also useful. By showing what journals cite what journals, and with what frequency, the *Journal Citation Reports* makes it possible to define the core and tail of the literature on any given subject, model journal networks, and to gauge the degree of interdisciplinary interaction in a proposed research project. Essentially, the data from *Journal Citation Reports* can be used to do all the things I have been talking about in regard to the management of science—the only difference being that the view is of science at the journal level of detail.

The citation-index view of the literature, then, extends deeply into the structure and dynamics of the scientific process itself. With the help of a computer, this view can be used to measure, define, and model the process at the level of individuals, papers, and journals. For those concerned with the study and management of science, that array of capabilities suggests some intriguing possibilities.

REFERENCES

1. **Aaronson, S.** "The Footnotes of Science." *Mosaic,* **6**:23-27, 1975.
2. **Martino, J.P.** "Citation Indexing for Research and Development Management." *IEEE Transactions on Engineering Management,* **EM-18**: 146-151, 1971.
3. **Sher, I. H., and Garfield, E.** "New Tools for Improving and Evaluating the Effectiveness of Research." In Yovits, M. C., Gilford, D.M., Wilcox R.H., Stavely, E., and Lemer H.D. L (eds.) *Research Program Effectiveness* (New York: Gordon and Breach, 1966), Pp. 135-146.
4. **Garfield, E.** "The 250 Most-Cited Primary Authors, 1961-1975. Part II. The Correlation between Citedness, Nobel Prizes, and Academy Memberships." *Current Contents,* No. 50: 5-16, December 12, 1977.

5. **Garfield, E.** "Citation Indexing for Studying Science." *Nature,* **227**:669–671, 1970.
6. **Clark, K.E.** "The APA Study of Psychologists." *American Psychologist,* **9**:117–120, 1954.
7. **Bayer, A.E. and Folger, J.** "Some Correlates of a Citation Measure of Productivity in Science." *Sociology of Education,* **39**:383–390, 1966.
8. **Orr, R.H. and Kassab, J.L.** "Peer Group Judgments on Scientific Merit: Editorial Refereeing." (Washington, D.C.: Congress of the International Federation of Documentation, October 15, 1965.)
9. **Virgo, J. A.** "A Statistical Procedure for Evaluating the Importance of Scientific Papers." *The Library Quarterly,* **47**:415–430, 1977.
10. **Price D.J.D.** "Networks of Scientific Papers." *Science,* **149**:510–515, 1967.
11. **Meadows, A.J.** "The Citation Characteristics of Astronomical Research Literature." *Journal of Documentation,* **23**:28–33, 1967.
12. **Wade, N.** "Citation Analysis: A New Tool for Science Administrators." *Science,* **188**:429–432, 1975.
13. **Geller, N.L., de Cani, J.S., and Davies, R. E.** "Lifetime Citation Rates as a Basis for Comparisons Within a Scientific Field." In Goldfield, E.D. (ed.) *American Statistical Association Proceedings of the Social Statistics Section.* (Washington, D.C.: American Statistical Association, 1975). Pp. 429–433.
14. **Price D.J.D.** "Citation Measures of Hard Science, Soft Science, Technology, and Non-Science." In Nelson, C.E. and Pollock, D.K. (eds). *Communication Among Scientists and Engineers* (Lexington, Mass.: D.C. Heath, 1970). Pp. 3–22.
15. **Garfield, E., Sher, I., and Torpie, R.J.** *The Use of Citation Data in Writing the History of Science* (Philadelphia: Institute for Scientific Information, 1964). 86 pp.
16. **Asimov, I.** *The Genetic Code* (New York: New American Library, 1962). 187 pp.
17. **Garfield, E., Malin, M.V., and Small, H.** "A System for Automatic Classification of Scientific Literature." *Journal of Indian Institute of Science,* **57**:61–74, 1975.

To Contents Page

To Chapter 7