To Contents Page

Chapter Seven

Citation Analysis as a Method of Historical Research into Science

The use of citation analysis in research on the history of science is based on a literary model of the scientific process. In this model, scientific work is represented by the papers written and published to report it, and the relationships between discrete pieces of work are represented by the references in the papers. Price, one of the leading contributors to the model, has taken this view of the scientific process to the point of defining scientific papers as the chief product of a scientist's work, and a scientist as one who writes scientific papers (1). Though the literary model is certainly a gross simplification of the scientific process, it seems to provide a functional view of that process that is both accurate and useful. Price has used it in a series of studies that have produced a number of insights into how science works and the ways in which it differs from, and interacts with, technology (1–4). Small and Griffith have used it to define the specialities that make up the leading edge of scientific development (5, 6), and I have used it to clarify the interactions between broad fields of research (7).

The accuracy and productivity of the model for historical studies was tested and proved in 1964 in a study conducted for the Air Force Office of Scientific Research (8). That study was concerned with determining whether citation analysis could be used to develop an accurate and useful network diagram of the cumulative research that led to a given scientific breakthrough.

The idea was not entirely a new one. Bernal had used the network diagram technique in 1953 to show the antecedents and consequences of Pasteur's discovery of molecular asymmetry but had not based it on citation analysis (9). The potential usefulness of references for historical research was suggested in 1955 (10). Then, in 1960, Dr. Gordon Allen put together the two ideas of references and diagrams with the illustration shown in Figure 7.1. A picture of the chronological and citation relationships among the papers in a bibliography on the staining of nucleic acids, the

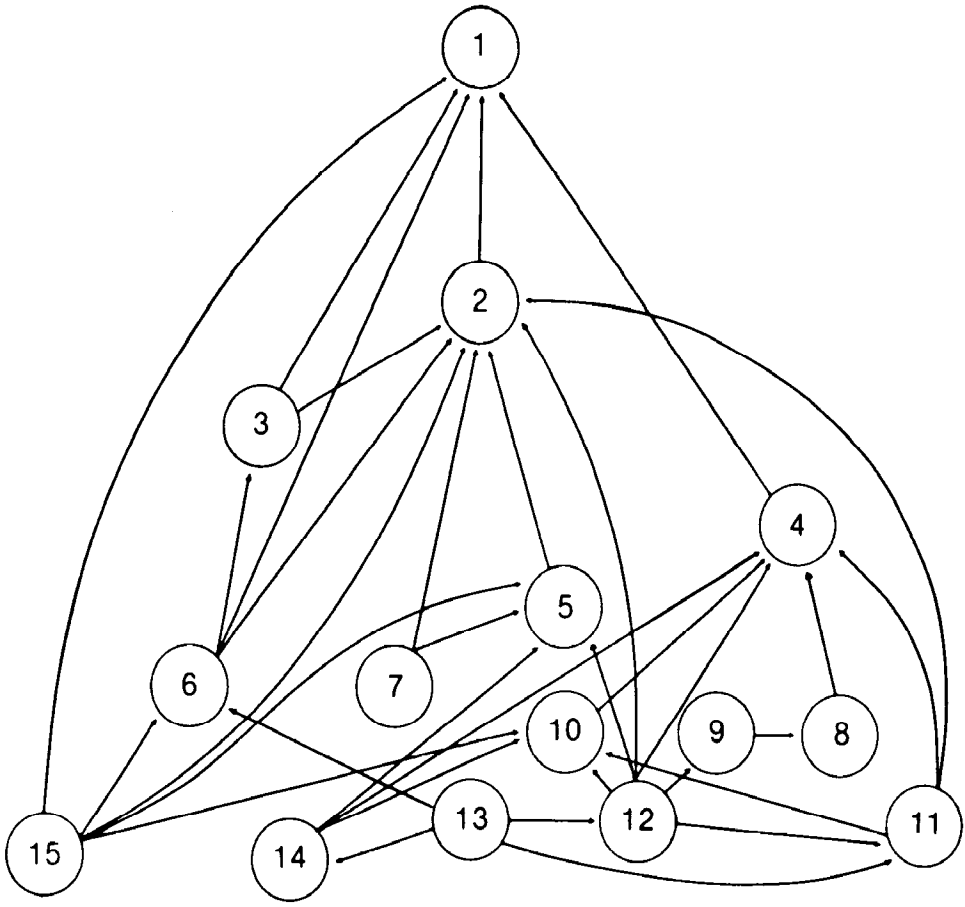


Figure 7.1 Citation network of the development of nucleic-acid staining.

KEY

1. Rabinowitch 1941
2. Michaelis 1947
3. Michaelis 1950
4. Zanker 1952
5. Northland 1954
6. Lawley 1956
7. Peacocke 1956
8. Appel 1958
9. Appel 1958
10. Steiner 1958
11. Steiner 1959
12. Bradley 1959
13. Bradley 1959
14. Bradley 1960
15. Loeser 1960

diagram uses circles arranged vertically in chronological order to represent papers and has arrows to represent the references between papers. Though Allen had not intended it, the resulting network diagram struck me as being a concise, easily understood outline of the historical development of the staining methodology. That observation led directly to the idea of using references to diagram the research dynamics of a given scientific development over time (11). According to the literary model of the scientific process, that type of analysis and presentation could be expected to be very useful to science historians. Bernal (12), Price (13), Leake (14), and Shryock (15) agreed that the idea had merit. The study for the Air Force was designed to determine how much.

ESTABLISHING A BASE LINE

To test the historical accuracy of citation analysis, we needed a base line—a recent scientific breakthrough whose history had been analyzed and documented by a recognized authority. We chose as our base line *The Genetic Code,* by Dr. Isaac Asimov (16), a clear, concise account of more than a century of complex research that led, eventually, to the development and validation of the DNA theory of genetic coding that controls protein synthesis.

The study strategy was simple: produce a network diagram of the events and relationships described by Asimov, produce a second diagram from the references in the papers that reported the Asimov events, compare the two to see how closely they match, and perform a thorough citation analysis of the papers to see whether they identify any important events or relationships missed by Asimov. How the strategy was implemented is shown in Figure 7.2.

The first task was to analyze Asmiov's account of the development to identify the

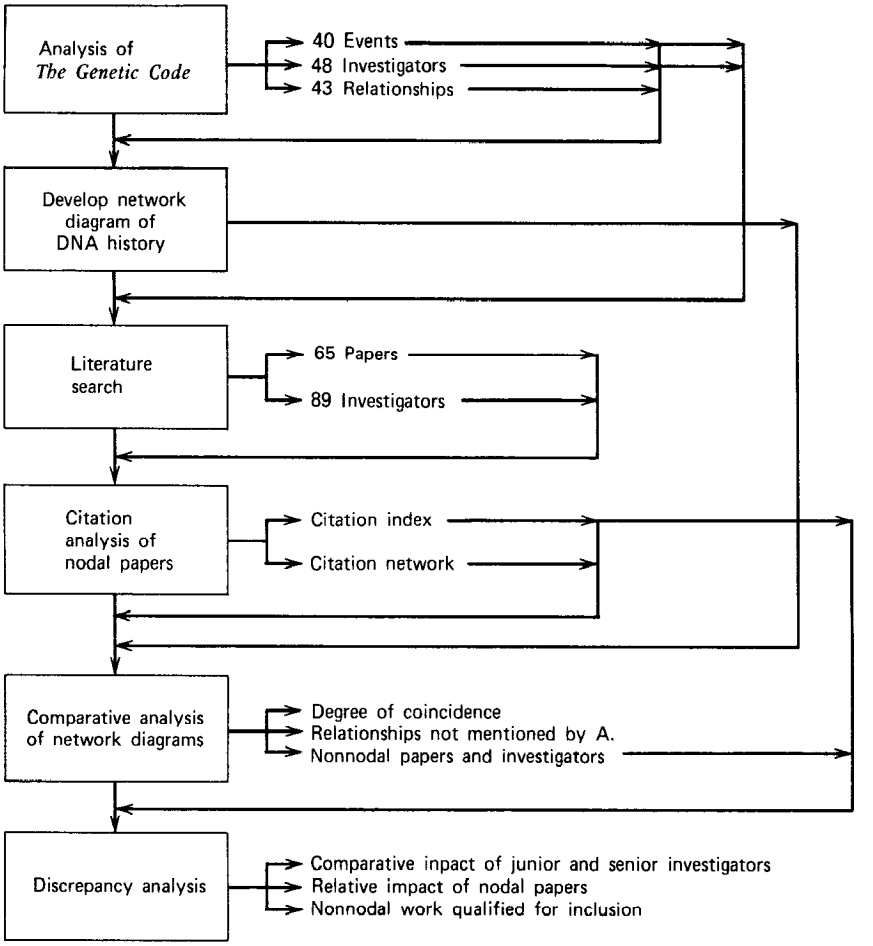


Figure 7.2 Flowchart of major tasks in study to validate use of citation analysis in defining the history of scientific development.

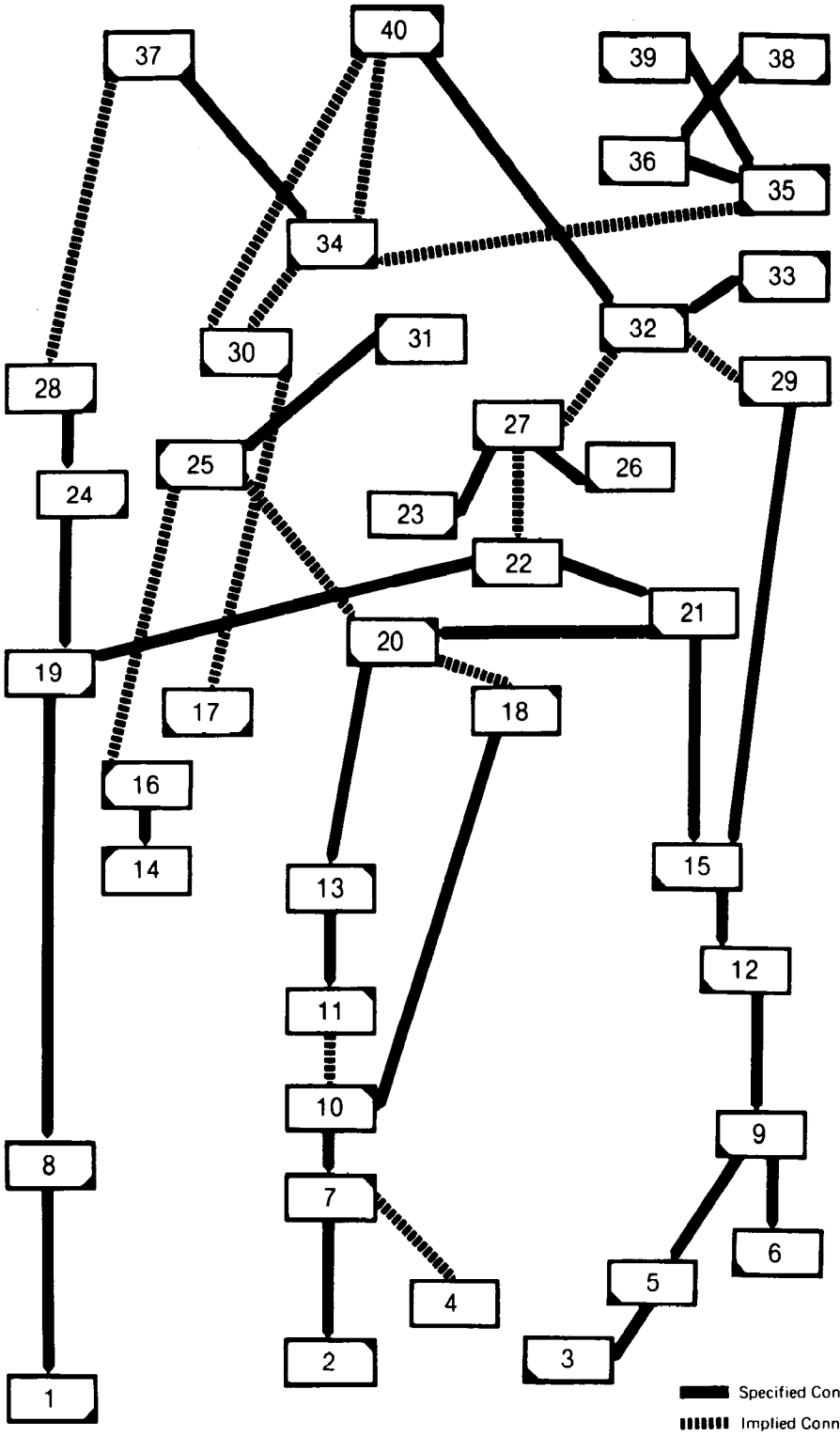


Figure 7.3 Network diagram of how DNA theory was developed and proved, as defined by I. Asimov in *The Genetic Code.*

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

research events and relationships he described. Forty events were found, ranging in time from 1820 to 1962. The descriptions of 36 of them included the names of the investigators involved; the remaining four did not, but they did give enough other information for us to be able to identify the investigators. Asimov also identified 29 relationships between events and implied the existence of another 14.

All 40 events and 43 relationships were then laid out in a network diagram (Figure 7.3) in which the nodes represent events and the arrows between nodes represent research relationships. Each node is numbered and identifies the name of the investigator credited with the research, the years covered by the research, and the general type of research. The type of research is shown by the corner code, which distinguishes between genetics, protein chemistry, nucleic acid chemistry, and virology. The nodes are grouped by type of research along three vertical lines to show the development and evolution of the three oldest research fronts. Protein chemistry events are on the left, nucleic acid chemistry on the right, and genetics in the middle. The diagram shows each of them as having been distinctly separate lines of research in the nineteenth century (bottom of diagram) and then combining to form molecular biology about the middle of this century (middle and top of diagram).

THE CITATION-BASED NETWORK

Developing a network diagram from the references of the papers that reported the nodal events began with an extensive literature search to determine which papers should be used. The search was conducted, on the names of investigators and sub-

jects obtained and derived from the Asimov book, in a number of major indexes, namely *Chemical Abstracts, Current List of Medical Literature,* and *Index Medicus.*

Though there was no problem in finding papers on the subjects of the nodal events, there was one in finding the particular papers that first reported the events as Asimov described them. The problem was one of judgment. Certain events were reported in a number of papers, all of which had to be analyzed thoroughly by subject specialists to determine which one was the first to report the particular research described by Asimov. Generally, the most difficult choices were posed by the events that took place after 1945, which is when scientists began the practice of publishing significant results in several journals concurrently and of publishing the results of multistage research a stage at a time. One paper chosen, for example, was the thirty-second of a series.

Strict adherence to the self-imposed rule that the papers selected be the first to report the events as Asimov defined them was an important convention in the methodology of the study. A number of events were first reported in papers with few references and later elaborated on in papers containing extensive bibliographies. The Watson and Crick discovery of the molecular configuration of DNA, for example, was announced in two articles, published in *Nature,* with minimal bibliographies. Within the year, however, Watson and Rich published a paper on the same subject, in the *Proceedings of the National Academy of Sciences U.S.,* that had a much more extensive bibliography. By always selecting representative papers on the basis of first announcement rather than length of bibliography, the study team made the test of the citation analysis technique considerably more rigorous than it might have been. As a result, the citation-based network that was developed is a demonstration not of how much historical detail citation analysis can define, but of whether it can define at least enough to be useful in the study of science history.

The literature search and review produced a total of 65 papers, which reported the nodal events, and 89 investigators, who were credited with authorship of the papers. These papers and investigators were the basis of the citation analysis that was performed to develop the alternative network diagram.

For the purposes of this analysis, a citation index (see excerpt in Figure 7.4) was developed from the 65 nodal papers. The primary entries in the index were the reference citations from the 65 papers. Under each entry were listed the nodal papers that cited it.

Using the nodal citation index to identify reference connections between papers, the network diagram shown in Figure 7.5 was developed and promptly named a "historiograph" ("historiogram" might have been more appropriate). The nodes are the Asimov events. The arrows connecting the nodes, however, reflect the relationships between events that are identified not by Asimov, but by the references in the nodal papers. Some of the connections are strong ones, consisting of references with a formal citation to another nodal paper, to a relevant nonnodal paper by a nodal author, or to a nonnodal paper by the citing author that, in turn, cites a nodal paper. In other words, all the strong connections are explicit references, though some are through an intermediate self-citation, to papers by nodal authors that are on the subject of nodal events. Other connections are weaker, consisting of

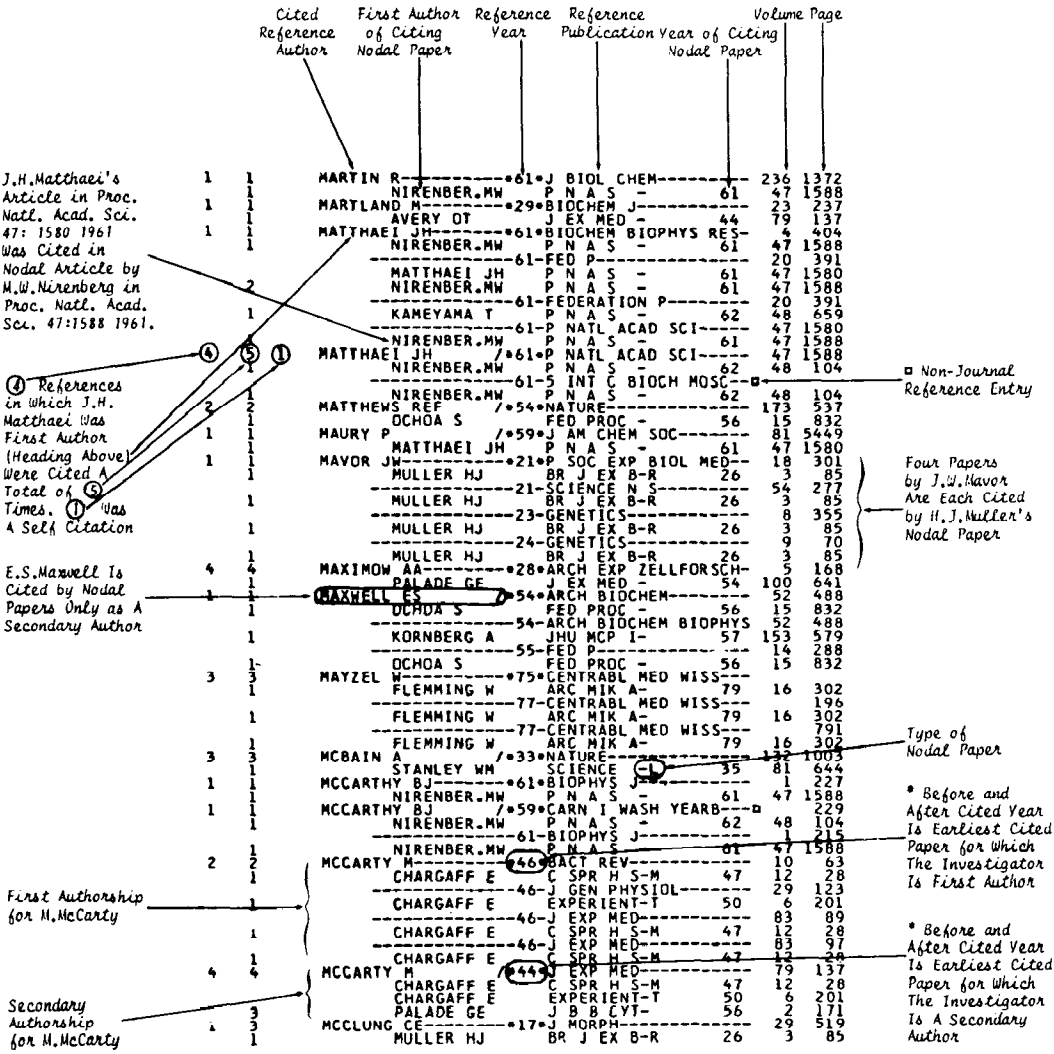


Figure 7.4 Excerpt of *Nodal Citation Index,* compiled from the bibliographies of the 65 papers that reported the events specified by I. Asimov in his history of the DNA theory.

acknowledgments of the relevant work of another nodal author without explicitly citing a particular paper, or references to papers by nonnodal authors that, in turn, cite a nodal paper.

The citation analysis performed to identify relationships was far from exhaustive. First of all, the nodal citation index was compiled only from the papers chosen to represent the nodal events. As mentioned earlier, these papers were only a fraction of what had been written to report some of the nodal events. If all the relevant papers by nodal authors had been included in the citation index, many more relationships between nodal events might have been identified. Second, the effort to uncover the relationships identified by even the limited citation index that was used concentrated primarily on looking for direct connections between nodal papers, regardless of whether the reference included a formal citation or stopped with an acknowledgment in the text. For economic reasons the search for indirect citations,

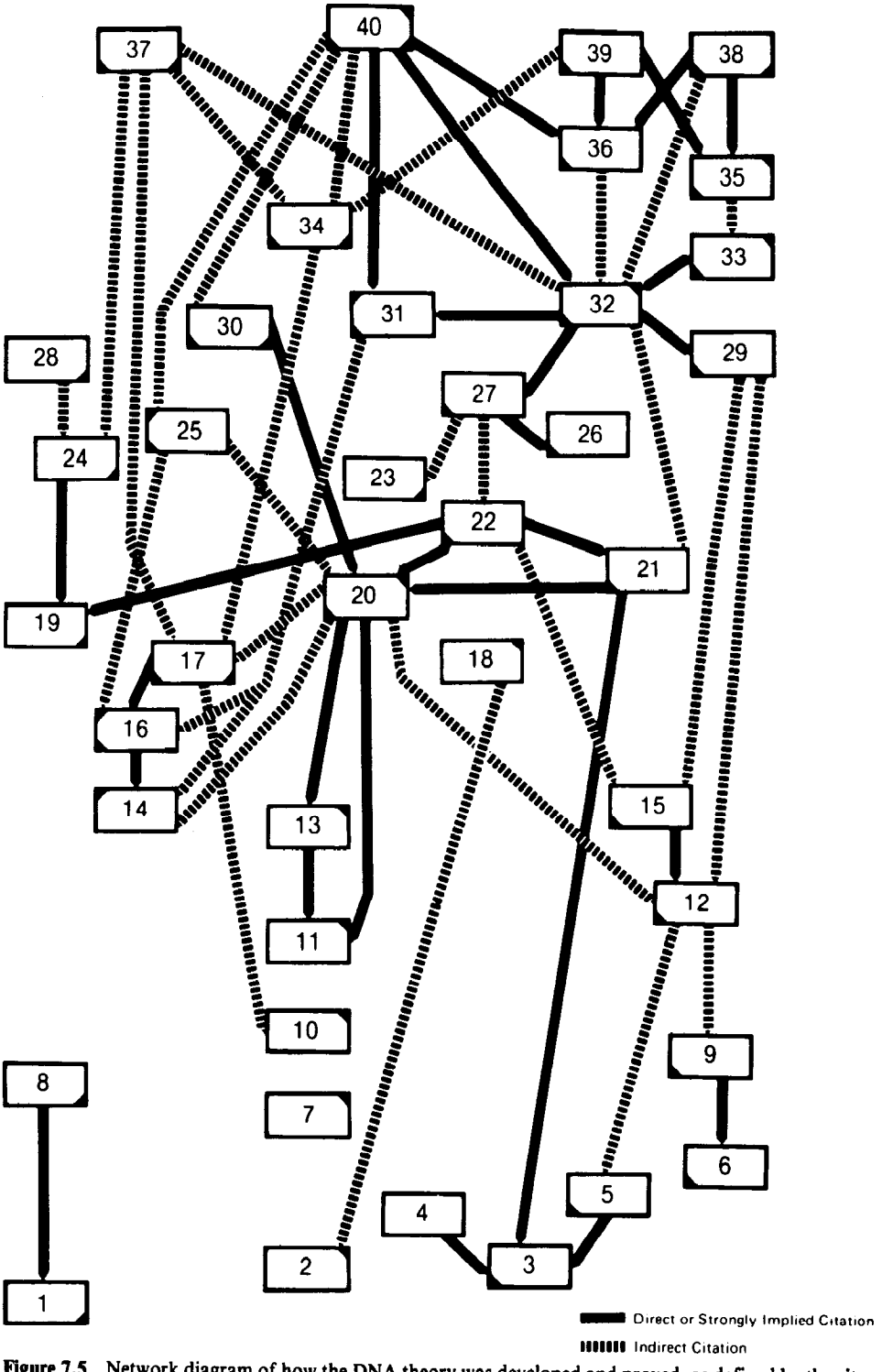


Figure 7.5 Network diagram of how the DNA theory was developed and proved, as defined by the citation connections among the nodal papers.

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

made through an intermediate reference, was conducted only in cases where no direct links existed. While this approach did not reduce the number of nodal events that could be linked together, it did understate the strength of some of the connections.

Despite these limitations, the citation analysis identified a total of 59 relationships between nodal events, and all but 11 of them were identified by strong reference links.

COMPARATIVE ANALYSIS

A comparative analysis of the historiographs produced from the Asimov account and the citation analysis turned up a number of similarities and differences. First, the similarities:

The most important similarity was that the historiograph produced from the citation analysis duplicated 65% of the relationships in the one produced from the Asimov account (28/43). The degree of coincidence was even greater for the relationships that Asimov considered important enough to specify in full detail. The citation-based historiograph duplicated 72% of them (21/29).

Two other points of similarity had to do with judgments made or implied about the relative originality and importance of the events. The citation historiograph showed 11 events that were not connected to any earlier work, which suggested that they were significant departures from earlier work, marked by an exceptional degree of originality, and probably of fundamental importance to the overall line of research. In some cases, the research independence is more apparent that real; it can be explained by the fact that the earlier a work appears on a chronological scale that

stretches back to the middle of the nineteenth century, the lower the probability that it will contain references to earlier work. Nevertheless, the citation historiograph's implication that these 11 events were historically independent coincided closely with Asimov's judgment. He related only four of them to earlier work, which means that his judgment confirms 64% of the inferences about fundamental importance that could be made from the citation historiograph. An analysis of the papers that reported the 11 events showed that all of them did, in fact, involve highly original work that opened up productive new directions.

The second similarity of judgment was more explicit. Asimov singled out one particular event as probably being the single most important contribution to the overall effort. To test the citation analysis against this judgment, we calculated citation weights for all the nodes. The weight assigned to each node reflected the number and type of reference links to and from all the other nodes in the network. The nodal event with the highest citation weight turned out to be the same one that Asimov had judged to be the most important.

The major point of difference between the two historiographs was that the one based on citation analysis identified 31 relationships not noted by Asimov. These relationships ranged in strength from perfunctory acknowledgment of earlier work to strong research dependency.

Other differences were apparent when the Asimov historiograph was compared with the nodal citation index, which identified a number of papers and investigators neither implied nor mentioned by Asimov. The papers were on work that did not correspond to any of the nodal events but that were important enough to have been cited by nodal papers. Some of the investigators Asimov failed to mention consisted of the authors of these nonnodal papers; the rest were uncredited coauthors of nodal papers. These points of discrepancy were analyzed in more detail to see if citation counts could be taken further as a measure of the impact of scientific work, and if citation analysis could identify any major contributions that Asimov had not.

DISCREPANCY ANALYSIS

The test of citation counts as a measure of impact was built around the 41 coauthors of nodal papers who Asimov did not credit. By not crediting them, he implied that their work had less impact than the work of those he did credit. The 1961 *SCI* was used to find out whether this implied difference was reflected in the citation record compiled that year for the two groups of investigators.

The citation record put together for each investigator was based on all papers cited in 1961 (not just nodal papers) and consisted of the following data:

1. Number of times cited.
2. Number of times cited by nonnodal authors.
3. Number of self-citations.
4. Number of times cited by the coauthors of their nodal papers.

5. Number of times cited by other nodal authors.
6. Publication date of the earliest paper cited.

Averages were then worked out for each class of investigators.

The analysis showed that the investigators credited by Asimov and designated as "senior" were more heavily cited than those who had not been credited by him and had been designated as "junior." In all but three cases, the junior investigators were cited less frequently than the senior investigator with whom they shared the authorship of a nodal paper. The three exceptions were all a matter of special circumstance: In one case, the junior and senior investigators were coauthors in a series of heavily cited papers, including the nodal paper, in which the junior investigator was listed more often as the first author. In another case, the junior investigator, again, was listed as first author on the nodal paper, which was a heavily cited one. And in the third case, the junior investigator had been publishing much longer than the senior one; when the papers published prior to the earliest cited paper of the senior investigator were excluded from the comparison, the senior investigator turned out to be cited more frequently.

The averages quantified the extent of the difference between the two groups. The 48 senior investigators each were cited an average of 112 times, while the average per junior investigator was only 41.6. These rates are put into perspective by a 5.5 average for all the reference authors listed in the 1961 *SCI* and a 169 average for the 13 winners of the Nobel Prize for physics, chemistry, and medicine in 1962 and 1963.

Since the analysis showed that citation counts did, in fact, reflect the type of gross judgment Asimov implied about the relative impact of scientific work, it was extended to see if citation rates could provide a more precise measure. This extension of the analysis was based on the assumption that the work reported in nodal papers probably had more impact than all, or most, other work reported by the authors. The question to be answered was whether the citation rates of the nodal papers reflected this level of quality.

Again, the 1961 *SCI* was the source of the citation data used. This time, however, the comparison was not between investigators, but between all the cited papers—nodal and nonnodal—in which each investigator was listed as the first author. The comparison was made by ranking each author's papers by the number of times they had been cited in 1961, and then examining the listing for each author to see where the nodal paper ranked relative to the others. A finding that a significant percentage of the nodal papers ranked first in their listings would indicate that citation counts might be a way of identifying what work by a given scientist has had the greatest impact.

There were a number of factors that made the results of this analysis more indicative than definitive. One was that the analysis was based on citation data taken from the single year of 1961. Because authors tend to cite recent literature more frequently than older material, this approach produced a bias that is inversely proportional to age. This tendency means that the older a paper is, the lower it is likely to

rank relative to more recent papers. The analysis confirmed the pattern when the average citation rate per nodal paper was computed for three different time periods: the average was 15.1 for articles published from 1951—1961, 5.5 for those published from 1930—1950, and only 1.1 for those published from 1820—1929. The total number of citations to all the papers in the earliest time period was so low, in fact, that there was doubt about their statistical significance.

Another factor distorting the results of the analysis was that the nodal papers were the earliest report of the Asimov events, but not always the most substantive. A number of authors went on to write comprehensive reviews of their nodal work that were much more heavily cited than their initial reports.

Then, too, it was reasonable to expect that some investigators would continue to do outstanding work that generated even more interest than their nodal research. This, of course, turned out to be the case in a few instances.

Despite these negative factors, the analysis showed that 61% of the nodal papers ranked as either the first or second most highly cited of the material published by their first author. As expected, the results were highly time sensitive. Only 35% of the nodal papers published prior to 1941 ranked as the first or second highest cited. But that ranking was achieved by 77% of the papers published after 1940. In fact, 54% of those published between 1941 and 1961 ranked as the most highly cited papers produced by the investigators listed as first authors.

The analysis concerned with identifying overlooked developments that qualified for inclusion in the Asimov account focused on the nonnodal papers and authors that were cited by at least three different nodes of the citation-based historiograph. They were identified by the nodal citation index.

Only one paper met that criterion. It also matched the primary citation characteristics of the nodal papers: its author had been cited in the 1961 *SCI* a total of 172 times, which was higher than the 112 citations averaged by the senior nodal investigators, and this particular paper was the most highly cited of his works. When the paper was reviewed, however, it was found to describe an experimental method that, though useful, probably was not important enough in the historical scheme of things to have been mentioned by Asimov.

The second step in the analysis dug a little deeper by identifying all the nonnodal authors who had been cited by at least three different nodes, but not for any one paper. This uncovered 26 investigators who merited additional study. Twenty-five of them were cited in the 1961 *SCI* more frequently than the average of 41.6 for the junior nodal authors. Thirteen of them were cited more frequently than the 112 average for the senior nodal investigators. And four of the 13 were cited by nodal authors for papers that ranked either first or second in citation rate relative to the rest of the reference citations listed for them in the 1961 *SCI*. Since these four papers matched the primary citation characteristics of the nodal papers, they were selected for individual analysis.

The individual analysis showed that two of the papers described methods, one a reaction, and one a phenomenon that provided an explanation of RNA replication in the absence of DNA. Though important, the methods and reaction probably were

not sufficiently critical to the development and proof of the DNA theory to have been mentioned by Asimov. The last paper, however, was a different story. RNA replication in the absence of DNA challenged the entire DNA theory. Work that brought that capability into the framework of the DNA theory certainly seemed to have a place in the history of the basic science involved. Asimov agreed with that conclusion in later discussions.

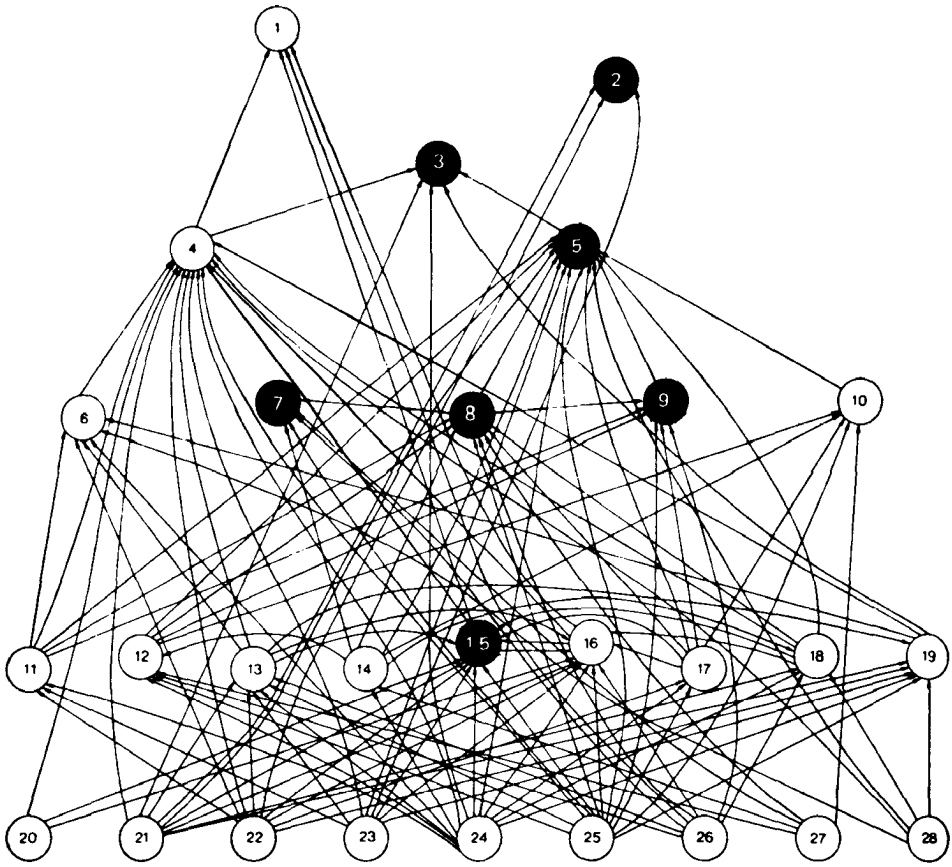
LESS EFFORT, MORE RESULTS

So the citation analysis ended up uncovering an event of importance that had been overlooked in a history written from memory by a scientist/writer whose memory is acknowledged to be phenomenal. Equally important, the study showed that citation analysis, even at a level considerably less than exhaustive, provides a way of identifying key events, their chronology, relationships, and relative importance, and that it is a very useful tool in working out the history of a given scientific effort. It could well be, in fact, that the technique is even more useful than the DNA study indicates. When the *SCI* data base has been expanded to cover the literature back as far as 1900, it will be possible to see whether the ability to analyze a much larger percentage of the important journal sources produces a significantly greater degree of historical definition.

Another aspect of the utility of citation analysis as a research method is that it is mechanistic. A computer can be used to compile a citation index from the bibliographic inputs, do much of the analysis needed to identify relationships and relative importance, and even produce historiographs (17). An on-line system for performing these functions in an interactive mode has already been developed at ISI (18). Even if the role of the computer is minimized to nothing more than the compilation of the index, the method is still mechanistic (algorithmic) in the sense that it consists of procedures that require no special knowledge or talents in either history or the subject being researched.

The historiograph in Figure 7.6 makes the point very nicely. An update of the DNA history that had been done in the 1964 study, it was produced by an assistant of mine, working under my direction, in 1968. Neither of us were historians, nor knew anything about genetics. Yet, it is quite accurate as far as it goes, which is to show the major advances in genetics since 1960.

The procedure we used was considerably simpler than the one followed in 1964 to test the validity of the method. Our starting point was a list of approximately three dozen papers obtained from a review of the 1967 literature on the subject (19). Using these papers as our source documents, we compiled a citation index of their bibliographies (20). The index consisted of several hundred papers, which we reduced to the 28 most important by the simple method of eliminating all that had been cited less than five times. To validate the importance of these papers, we checked them out in the 1967 *SCI* to see whether their relatively high citation rate was maintained within the broader framework of the approximately 304,000 source



1, Sheehan 1958; 2, Bray 1960; 3, Nirenberg 1961; 4, Marcker 1964; 5, Nirenberg 1964; 6, Marcker 1965; 7, Brenner 1965; 8, Khorana 1965; 9, Nirenberg 1965; 10, Khorana 1965; 11, Marcker 1966; 12, Khorana 1966; 13, Marcker 1966; 14, Khorana 1966; 15, Adams 1966; 16, Webster 1966; 17, Nirenberg 1966; 18, Ochoa 1966; 19, Nakamoto 1966; 20, Berberich 1967; 21, Lucas-Leonard 1967; 22, Caskey 1967; 23, Ochoa 1967; 24, Khorana 1967; 25, Nirenberg 1967; 26, Ochoa 1967; 27, Khorana 1967; 28, Ochoa 1967.

Figure 7.6 Historiograph of the major advances in genetics between 1958 and 1967, based on a citation analysis of a review of the 1967 literature. Each circle represents a paper cited five or more times by the papers listed in the bibliography of the review. The papers represented by solid black circles were cited 15 times or more in the 1967 *SCI*.

items from which that edition of *SCI* was compiled. They were, and the citation index was used to draw the historiograph shown in Figure 7.6. A bibliography of its nodal papers is shown in Figure 7.7.

Spanning the 10-year period from 1958 through 1967, this historiograph certainly does not fill in completely everything that happened since the earlier study, but it does provide a useful outline of the core work done in that time period. If the analysis had been expanded to the literature of each of the intervening years between the two studies, the picture would have been proportionately more comprehensive.

node

1. SHEEHAN, J. C. and YANG, D. M. (1958), The use of N-formylamino acids in peptide synthesis. *J. Amer. Chem. Soc.*, 80, 1154.
2. BRAY, G. A. (1960), A simple efficient liquid scintillator for counting acqueous solutions in a liquid scintillation counter. *Analyt. Biochem.*, 1, 279.
3. NIRENBERG, M. and MATTHAEI, J. H. (1961), The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc. nat. Acad. Sci. (Wash.)*, 47, 1588.
4. MARCKER, K. A. and SANGER, F. (1964), N-formylmethionyl-sRNA. *J. molec. Biol.*, 8, 835.
5. NIRENBERG, M. and LEDER, P. (1964), RNA codewords and protein synthesis-effect of trinucleotides upon binding of sRNA to ribosomes. *Science*, 145, 1399.
6. MARCKER, K. (1965), Formation of N-formyl-methionyl-sRNA. *J. molec. Biol.*, 14, 63.
7. BRENNER, S., STRETTON, A. O. W. and KAPLAN, S. (1965), Genetic code — nonsense triplets for chain termination and their suppression. *Nature*, 206, 994.
8. SÖLL, D., OHTSUKA, E., JONES, D. S., LOHRMANN, R., HAYATSU, H., NISHIMURA, S. and KHORANA, H. G. (1965), Studies on polynucleotides. 49. Stimulation of binding of aminoacyl-SRNAS to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids. *Proc. nat. Acad. Sci. (Wash.)*, 54, 1378.
9. NIRENBERG, M., LEDER, P., BERNFIELD, M., BRIMACOMBE, R., TRUPIN, J., ROTTMAN, F. and O'NEAL, C. (1965), RNA codewords and protein synthesis. 7. On general nature of RNA code. *Proc. nat. Acad. Sci. (Wash.)*, 53, 1161.
10. NISHIMURA, S., JONES, D. S., OHTSUKA, E., HAYATSU, H., JACOB, T. M. and KHORANA, H. G. (1965), Studies on polynucleotides. 47. *In vitro* synthesis of homopeptides as directed by a ribopolynucleotide containing a repeating trinucleotide sequence — new codon sequences of lysine glutamic acid and arginine. *J. molec. Biol.*, 13, 283.
11. BRETSCHER, M. S. and MARCKER, K. A. (1966), Polypetidyl-s-ribonucleic acid and aminoacyl-s-ribonucleic acid binding sites on ribosomes. *Nature*, 211, 380.
12. JONES, D. S., NISHIMURA, S. and KHORANA, H. G. (1966), Studies on polynucleotides. 56. Further syntheses in vitro of copolypeptides containing 2 amino acids in alternating sequence dependent upon DNA-like polymers containing 2 nucleotides in alternating sequence. *J. molec. Biol.*, 16, 454.
13. CLARK, B. F. C. and MARCKER, K. A. (1966), N-formyl-methionyl-s-ribonucleic acid and chain initiation in protein biosynthesis — polypeptide synthesis directed by a bacteriophage ribonucleic acid in a cell-free system. *Nature*, 211, 378.
14. MORGAN, A. R., WELLS, R. D. and KHORANA, H. G. (1966), Studies on polynucleotides. 59. Further codon assignments from amino acid incorporations directed by ribopolynucleotides containing repeating trinucleotide sequences. *Proc. nat. Acad. Sci. (Wash.)*, 56, 1899.
15. ADAMS, J. M. and CAPECCHI, M. R. (1966), N-formylmethionyl-sRNA as initiator of protein synthesis. *Proc. nat. Acad. Sci. (Wash.)*, 55, 147.
16. WEBSTER, R. E., ENGELHARDT, D. L. and ZINDER, N. (1966), In vitro protein synthesis — chain initiation. *Proc. nat. Acad. Sci. (Wash.)*, 55, 155.
17. KELLOGG, D. A., DOCTOR, B. P., LOEBEL, J. E. and NIRENBERG, M. (1966), RNA codons and protein synthesis. 9. Synonym codon recognition by multiple species of valine-, alanine-, and methionine-sRNA. *Proc. nat. Acad. Sci. (Wash.)*, 55, 912.
18. STANLEY, W. M., SALAS, M., WAHBA, A. J. and OCHOA, S. (1966), Translation of genetic message — factors in initiation of protein synthesis. *Proc. nat. Acad. Sci. (Wash.)*, 56, 290.
19. NAKAMOTO, T. and KOLAKOFSKY, D. (1966), A possible mechanism for initiation of protein synthesis. *Proc. nat. Acad. Sci. (Wash.)*, 55, 606.
20. BERBERICH, M. A., KOVACH, J. S. and GOLDBERGER, R. F. (1967), Chain initiation in a polycistronic message — sequential versus simultaneous derepression of enzymes for histidine biosynthesis in Salmonella typhimurium. *Proc. nat. Acad. Sci. (Wash.)*, 57, 1857.
21. LUCAS-LENARD, J. and LIPMANN, F. (1967), Initiation of polyphenylalanine synthesis by N-acetylphenylalanyl/sRNA. *Proc. nat. Acad. Sci. (Wash.)*, 57, 1050.
22. CASKEY, C. T., REDFIELD, B. and WEISSBACH, H. (1967), Formylation of guinea pig liver methionyl-sRNA. *Arch. Biochem.*, 120, 119.
23. SALAS, M., HILLE, M. B., LAST, J. A., WAHBA, A. J. and OCHOA, S. (1967), Translation of genetic message. 2. Effect of initiation factors on binding of formyl-methionyl-tRNA to ribosomes. *Proc. nat. Acad. Sci. (Wash.)*, 57, 387.
24. GHOSH, H. P., SÖLL, D. and KHORANA, H. G. (1967), Studies on polynucleotides. 67. Initiation of protein synthesis in vitro as studied by using ribopolynucleotides with repeating nucleotide sequences as messengers. *J. molec. Biol.*, 25, 275.
25. MARSHALL, R. E., CASKEY, C. T. and NIRENBERG, M. (1967), Fine structure of RNA codewords recognized by bacterial amphibian and mammalian transfer RNA. *Science*, 155, 820.
26. LAST, J. A., STANLEY, W. M., SALAS, M., HILLE, M. B., WAHBA, A. J. and OCHOA, S. (1967), Translation of genetic message. 4. UAA as a chain termination codon. *Proc. nat. Acad. Sci. (Wash.)*, 57, 1062.
27. KÖSSEL, H., MORGAN, A. R. and KHORANA, H. G. (1967), Studies of polynucleotides. 73. Synthesis in vitro of polypeptides containing repeating tetrapeptide sequences dependent upon DNA-like polymers containing repeating tetranucleotide sequences — direction of reading of messenger RNA. *J. molec. Biol.*, 26, 449.
28. SALAS, M., MILLER, M. J., WAHBA, A. J. and OCHOA, S. (1967), Translation of genetic message. 5. Effect of Mg^{++} and formylation of methionine in protein synthesis. *Proc. nat. Acad. Sci. (Wash.)*, 57, 1865.

Figure 7.7 Bibliography of the nodal papers in the historiograph of the major advances in genetics between 1958 and 1967.

Citation analysis, then, seems to be a method that greatly simplifies the effort involved in constructing the sequence of events and web of relationships that serve as the starting point for the evaluations, interpretations, and explanations that are the essence of historical research.

There is, of course, one factor that limits the application of citation analysis in history-of-science studies. Bibliographic citation has been an established convention of scientific publication only since the early part of the twentieth century. The further back in time a study goes beyond that point, the less realistic the picture produced by citation analysis. In historical studies dealing with developments since the first quarter of this century, however, citation analysis is a method that seems to be able to simplify the research process and increase the research results.

REFERENCES

1. **Price, D.J.D.** "Is Technology Historically Independent of Science? A Study in Statistical Historiography." *Technology & Culture,* 1:553–568, 1965.
2. **Price, D.J.D.** "Networks of Scientific Papers." *Science,* 149:510–515, 1967.
3. **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.
4. **Price, D.J.D.** *Little Science, Big Science* (New York: Columbia University Press, 1963). 119 pp.
5. **Small, H.G. and Griffith, B.C.** "The Structure of Scientific Literature, I: Identifying and Graphing Specialties." *Science Studies,* 4:17–40, 1974.
6. **Griffith, B.C., Small, H.G., Stonehill, J.A., and Dey, S.** "The Structure of Scientific Literatures, II: Toward a Macro- and Micro-structure for Science." *Science Studies,* 4:339–365, 1974.
7. **Garfield, E.** "Journal Citation Studies 1: What is the Core Literature of Biochemistry as Compared to the Core of Chemistry?" *Essays of an Information Scientist,* Vol. 1 (Philadelphia: ISI Press, 1977). Pp. 262–265.

   Garfield, E. "Journal Citation Studies 2: What is the Core Literature of Chemical Physics?" *Essays of an Information Scientist,* Vol. 1 (Philadelphia: ISI Press, 1977). Pp. 274–277.

   Garfield, E. "Journal Citation Studies 3: *Journal of Experimental Medicine* Compared With *Journal of Immunology,* or How Much of a Clinician is the Immunologist?" *Essays of an Information Scientist,* Vol. 1 (Philadelphia: ISI Press, 1977). Pp. 326–329

   Garfield, E. "Journal Citation Studies 4: The Literature Cited in Rheumatology is Not Much Different From That of Other Specialties." *Essays of an Information Scientist,* Vol. 1 (Philadelphia: ISI Press, 1977). Pp. 338–341.

   Garfield, E. "Journal Citation Studies 5: Is Paleontology a Life or a Physical Science? *JCI* Reveals Gap in Coverage of Paleontology and Need for Better Small Journal Statistics." *Essays of an Information Scientist,* Vol. 1 (Philadelphia: ISI Press, 1977). Pp. 423–424.

   Garfield E. "Journal Citation Studies 6: *Journal of Clinical* Investigation: How Much Clinical and How Much Investigation?" *Essays of an Information Scientist,* Vol. 2 (Philadelphia: ISI Press, 1977). Pp. 13–16.

   Garfield, E. "Journal Citation Studies 10: Geology and Geophysics." *Essays of an Information Scientist,* Vol. 2 (Philadelphia: ISI Press, 1977). Pp. 102–106.

   Garfield, E. "Journal Citation Studies 14: Wherein We Observe That Physicists Cite Different Physics Journals Than Other People." *Essays of an Information Scientist,* Vol. 2 (Philadelphia: ISI Press, 1977). Pp. 154–157.

Garfield, E. "Journal Citation Studies 15: Cancer Journals and Articles." *Essays of an Information Scientist,* Vol. 2 (Philadelphia: ISI Press, 1977). Pp. 160–167.

Garfield, E. "Journal Citation Studies 19: Psychology and Behavior Journals." *Essays of an Information Scientist,* Vol. 2 (Philadelphia: ISI Press, 1977). Pp. 231–235

Garfield, E. "Journal Citation Studies 20: Agriculture Journals and the Agricultural Literature." *Essays of an Information Scientist,* Vol. 2 (Philadelphia: ISI Press, 1977). Pp. 272–278.

Garfield, E. "Journal Citation Studies 21: Engineering Journals." *Essays of an Information Scientist,* Vol. 2 (Philadelphia: ISI Press, 1977). Pp. 304–309.

8. **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.
9. **Bernal, J.D.** *Science and Industry in the Nineteenth Century* (London: Routledge, Kegan Paul Ltd., 1953), P. 23.
10. **Garfield, E.** "Citation Indexes for Science." *Science,* **122**:108–111, 1955.
11. **Garfield, E.** "Citation Indexes in Sociological and Historical Research." *American Documentation,* **14**:289–291, 1963.
12. **Bernal, J.D.** Private communication, March 1962.
13. **Price, D.J.D.** Private communication, March 1962.
14. **Leake, C.D.** Private communication, August 1962.
15. **Shryock, R.** Private conversation, September 1962.
16. **Asimov, I.** *The Genetic Code* (New York: New American Library, 1963). 187 pp.
17. **Garfield, E., and Sher, I.H.** *Diagonal Display—A New Technique for Graphic Representation of Complex Topological Networks* (Philadelphia: Institute for Scientific Information, 1967). 94 pp.
18. **Yermish, I.** *"A Citation-Based Interactive Associative Information Retrieval System."* (Philadelphia: University of Pennsylvania, Ph.D. dissertation, 1975). 278 pp. mimeogr.
19. **Sadgopal, A.** "Genetic Code After the Excitement." *Advances in Genetics,* **14**:325–404.
20. **Garfield, E.** "Citation Indexing, Historio-Bibliography, and the Sociology of Science." *Proceedings of the Third International Congress of Medical Librarianship,* Amsterdam, May 1969, Davis, K.E. and Sweeney, W.D. (eds.) (Amsterdam: Excerpta Medica, 1970). pp. 187–204. Reprinted in: **Garfield, E.** *Essays of an Information Scientist,* Vol. 1 (Philadelphia: ISI Press, 1977). Pp. 158–174.

To Contents Page

To Chapter 8