#### THE USE OF CITATION DATA IN WRITING THE HISTORY OF SCIENCE

December 31, 1964

Eugene Garfield, Ph.D., *Director* Irving H. Sher, Sc.D., *Director of Research* Richard J. Torpie, *Research Associate* 

© Copyright 1964 by the Institute for Scientific Information Inc. Philadelphia, Pennsylvania, USA

All rights reserved. This book, or parts thereof, may not be reproduced in any form without permission of the publisher.

Library of Congress Catalog Card Number 64-66453

Printed in the United States of America

#### ACKNOWLEDGEMENT

We wish to thank Dr. Isaac Asimov for his willingness to submit his book to the liberties of our discussion. Our appreciation is extended to his publisher, the New American Library, for permission to prepare a synopsis of the text. Mrs. Diane Ippoldo is to be complimented for her patient artwork in preparing the network diagrams.

This work constitutes a final report of research for the Air Force Office of Scientific Research under Contract AF 49(638)-1256. DD Form 1473. See last page.

#### TABLE OF CONTENTS

REFERENCES CITED	<ul> <li>ii</li> <li>iii</li> <li>1</li> <li>3</li> <li>7</li> </ul>							
SUMMARY       . </td <td><pre>iii    1    3    7</pre></td>	<pre>iii    1    3    7</pre>							
INTRODUCTION	1 3 7							
METHODOLOGY	3 7							
ANALYSIS OF THE CITATIONS TO NODAL AUTHORS FROM THE 1961 SCIENCE CITATION INDEX	7							
1961 SCIENCE CITATION INDEX	7							
TABLE 1 1961 SCIENCE CITATION INDEX CITATIONS TO								
NODAL AUTHORS	8							
A. Comparison of Senior and Junior Nodal Authors	11							
B. Breakdown of the Total Count by Type of Citation	11							
1. Senior Nodal Author	11							
2. Junior Nodal Author	11							
3. Self Citations	11							
4. Citation By Coauthor	12							
5. Citation By Other Nodal Authors	12							
C. Retrospect: The 1961 Citation of a Nodal Author by the Author of a								
Different Node	12							
D. Citation "Leapfrogging" Effect	13							
E. Chronological Position: An Analysis of the Earliest Cited Work by a Nodal Author	14							
Histogram: Citation Leapfrogging by Nodes (Fig. i)	14							
ANALYSIS OF THE CITATIONS TO NODAL ARTICLES FROM THE								
1961 SCIENCE CITATION INDEX	15							
A. Selection of the Nodal Article	15							
B. Ranking of Citation Counts to the Nodal Article	15							
Table 2: Ranking of Nodal Articles Relative to Other Cited         Works by the Same First Author Based on Citation         Counts Found In 1061 (or 1064) Science Citation Index	16							
Table 3: Chronological Summary of Table 2	10							
Table 4: Breakdown of all 65 Nodal Articles 1810-1962	10							
Table 5. Breakdown of the Most Decent 44 Nodel Articles	19							
1941-1961	19							
Table 6A: Citation Ranking of Pooled Nodal Papers for 41 NodalFirst Authors 1819-1961	20							
Table 6B: 1941-1961	20							
Table 6C: 1819-1941	20							
	<ul> <li>Different Node</li> <li>D. Citation "Leapfrogging" Effect</li> <li>Chronological Position: An Analysis of the Earliest Cited Work by a Nodal Author</li> <li>Histogram: Citation Leapfrogging by Nodes (Fig. i)</li> <li>ANALYSIS OF THE CITATIONS TO NODAL ARTICLES FROM THE</li> <li>1961 SCIENCE CITATION INDEX</li> <li>A. Selection of the Nodal Article</li> <li>B. Ranking of Citation Counts to the Nodal Article</li> <li>Table 2: Ranking of Nodal Articles Relative to Other Cited Works by the Same First Author Based on Citation Counts Found In 1961 (or 1964) Science Citation Index</li> <li>Table 3: Chronological Summary of Table 2</li> <li>Table 4: Breakdown of all 65 Nodal Articles 1819-1962</li> <li>Table 5: Breakdown of the Most Recent 44 Nodal Articles 1941-1961</li> <li>Table 6A: Citation Ranking of Pooled Nodal Papers for 41 Nodal First Authors 1819-1961</li> <li>Table 6B: 1941-1961</li> </ul>							

VII.	DIS	SCU	SSION OF THE CITATION INDEX PREPARED FROM THE SIXTY-	
	FI	/E	NODAL PAPERS (NCI)	23
	А.	Th	ne Nodal Citation Index (NCI) as a Method of Historical Investigation .	23
		1.	Selection of Articles Cited by at Least Three Separate Nodes	23
			Table 7: NCI and 1961 SCI Citation Analyses for Non-Nodal	
			Authors Cited by at Least Three Different Nodes	24
		2.	Selection of Non-Nodal Authors Cited by at Least Three Separate	
			Nodes	25
			a. Comparisons to Nodal Authors	25
			Table 8: The Number of Different Nodes Involved at Least	
			Once in the Citation of a Senior Nodal Author	25
			b. Selection of Potential Nodal Articles	26
			c. Evaluation of Potential Nodes	27
		3.	Coupling of Nodal Articles as Demonstrated in the NCI	28
		4.	Intermediate References Used in Indirect Citation Connections.	28
	В.	Hi	storical Network Chart	28
	C.	La	ck of Early Citation Dependency and Scientific Originality	29
VIII.	CO	NC	LUSIONS	31
APPE	ND	ICE	ES	
Ι	SY	NOI	PSIS OF THE BOOK, "THE GENETIC CODE" BY ISAAC ASIMOV .	35
II	DET	ΓAΙ	LED DESCRIPTION OF NODAL CITATION CONNECTIONS AND	
	WE	IGI	HTINGS IN THE NETWORK CHARTS	41
III	CIT	ГАТ	TION INDEX PREPARED FROM THE 65 NODAL PAPERS (NCI)	51
IV	WC	ORK	LOCATIONS SPECIFIED BY NODAL ARTICLES	58
V	AG	EN	CIES SUPPORTING THE RESEARCH	59
VI	SO	UR	CE INDEX OF NODAL PAPERS	60
VII	LE	GEI	ND FOR THE NETWORK CHARTS	74
	DD	14′	73 FORM	76

#### NETWORK CHARTS: FOLLOWING LAST PAGE

- 1. (Red) First Overlay..... Asimov's Specified Historical Connections
- 2. (Red) Second Overlay---Asimov's Implied Historical Connections
- 3. (Blue) Third Overlay .... Coincident Strong Citation Connections
- 4. (Blue) Fourth Overlay---Coincident Weak Citation Connections
- 5. (Yellow) Fifth Overlay--Non-Coincident Strong Citation Connections
- 6. (Yellow) Sixth Overlay--Non-Coincident Weak Citation Connections

#### I. FOREWORD

Can a computer write the history of science? Probably not in the sense usually implied. However, the research reported herein is a preliminary attempt to understand and define some basic problems that must be solved if computers are ever to aid the historian of science -- no less supplant him. In this study it was necessary to select a recent important scientific breakthrough which was based on the cumulation of years of diverse scientific achievement. For this reason we selected the discovery of the DNA code. For a concise historical description of the events, we then selected "The Genetic Code," a book by Dr. Isaac Asimov which describes the major scientific developments that eventually led to the duplication in the laboratory of the process of protein synthesis under control of DNA.

The choice of the genetic code as our case study was not fortuitous. Major breakthroughs in the field of molecular biology occurred at a time which coincided with the completion of our first extensive experimental citation indexes, the *Genetics Citation Index* (I) and the *1961 Science Citation Index* (2) from which part of the GCI was extracted. The availability of pertinent citation data made practical the testing of citation indexing for constructing historical maps and evaluating individual scientific events.

The history of citation indexing for the purposes of disseminating and retrieving information has been extensively described elsewhere (3). A suggestion for its use in historical research came as early as 1955 (4,5). However, the use of citation data for constructing historical maps was given great impetus by Dr. Gordon Allen when he prepared a bibliographic citation network diagram demonstrating the chronological relationship and citational linkages among a group of papers on the staining of nucleic acids. Allen's citation network diagram provided a useful model of scientific literature and simultaneously provided, in a two-dimensional topological display,- the historical development of the subject matter covered by the fifteen papers in his bibliography. (6) The availability of large files of computer-generated citation indexes and the experience derived in their preparation made practical the possibility of testing the usefulness of this approach in studying history.

The methodology developed here will hopefully prove useful to the historian and others interested in tracing the origins of discovery and creativity. It consisted of two steps.

First, we carefully identified the specific papers involved in the discoveries described by Asimov in his history of DNA. The exacting work in tracing all the pertinent citations should be readily apparent from examining the report. From this data we constructed a topological network diagram for 40 milestone events as described by Asimov. Then, we constructed a similar topological network based on citation data appearing in the bibliographies included in the papers reporting the same key discoveries.

The two networks were extensively analyzed and compared and demonstrated a high degree of coincidence between an historian's account of events and the citational relationship between these events. Comparison of the resulting networks has been facilitated by the use of special transparent overlays.

We also created a special citation index file from the references given in the papers reporting the milestone events described by Asimov. We elaborated on this basic corpus of citation data by drawing upon our broader *1961 Science Citation Index*.

Though this study was undertaken to investigate and test new methodologies for facilitating the writing of the history of science, we do not wish in any way to imply that the role of the scholar can be eliminated. The citation network technique does provide the scholar with a new *modus operandi* which, we believe, could and probably will significantly affect future historiography.

With the accelerating pace and complexity of scientific developments, the study of the history of science, research administration, and the sociology of science, now more than ever, can profitably employ new techniques for sifting and evaluting data. We believe the techniques described here can be of great utility for the administration of large-scale programs of research as well as for sociological and historical research.

#### **REFERENCES CITED**

- (1) E. Garfield and I.H. Sher, *Genetics Citation Index*, Institute for Scientific Information, Philadelphia, 1963, 864 pp.
- (2) E. Garfield and I.H. Sher, *1961 Science Citation Index*, Institute for Scientific Information, Philadelphia, 1963, 2672 pp.
- (3) E. Garfield, "Science Citation Index, A New Dimension in Indexing," Science 144(3619), 649-654 (1964). (See also introduction to the 1964 Science Citation Index).
- (4) E. Garfield, "Citation Indexes for Science," Science 122(3159), 108-111 (1955).
- (5) E. Garfield, "Citation Indexes in Sociological and Historical Research," *American Documentation* 14(4), 289-291 (1963).
- (6) G. Allen, Private Communication, June 2, 1960.
- (7) B.L. Clarke, Multiple Authorship Trends in Scientific Papers, Science 143:822 (1964).

#### II. SUMMARY

Writing the history of science has traditionally been a purely intellectual or cerebral pursuit of the scholar. A project is described herein which poses, and provides the first step toward the ultimate answer to the question "Can historical analysis be performed by a computer?" The more immediate goal was to test the initial hypothesis that citation indexes are useful heuristic tools for the historian. In this approach the history of science is regarded as a chronological sequence of events in which each new discovery is dependent upon earlier discoveries. Models of history are constructed consisting of chronologic maps or topological network diagrams. Two such models were used here. The first is based on the events in the history of DNA as described by Dr. Isaac Asimov in The *Genetic* Code. The second is based on the bibliographic citation data contained in the documents which are the origional published studies of events represented in the Asimov book. The interdependencies of linkages among 40 major events (nodes) included in both network diagrams were carefully mapped and compared.

A novel method was devised for these comparisons. Colored transparencies of the network diagrams, when superimposed, aid in the identification of historical dependencies between events. The red transparencies show those dependencies revealed by the Asimov analysis alone; the yellow transparencies show those dependencies revealed by citation data alone, and the blue transparencies show the dependencies commonto both analyses. Connecting lines between nodes were coded to indicate whether the linkages are explicit (in the case of Asimov) and direct or indirect (in the case of citations.)

The analyses, supported by numerous statistical tables and specially constructed citation indexes, show that the original hypothesis is reasonable. Unquestionably, bibliographic citation data, if presented in the form of network diagrams and or citation indexes, reveal historical dependencies which can be easily overlooked by the historian. On the other hand, citation standards are not always sufficiently rigorous to eliminate the need for human memory and evaluation, it is reasonable to conclude that the techniques described in this study can be profitably used in writing the history of science by helping to identify key events, their chronology, their interrelationships, and their relative importance.

In this study we first.carefully searched the scientific literature in order to determine the published works which most accurately fit each historical event described by Dr. Asimov. Altogether there were 65 "nodal" articles selected which had been written by 89 different investigators, 48 of whom are explicitly mentioned in Asimov's text. The 40 events, each of which is a node in the historical graph, were categorized and coded in broad subject classifications and arranged chronologically on transparent overlays. To determine citation linkages between nodes, the bibliographies of all nodal articles were first examined for direct citation to other nodal articles. Less direct citation linkages were also established through chronologically intermediate works by nodal authors, or in a few cases, where these were lacking, through intermediate works by non, nodal authors.

In this study, 65% (28/43) of the historical dependencies in the Asimov network were confirmed by corresponding linkages established by citations, In addition 31 citation connections were found which do not correspond to any historical dependencies noted in *"The Genetic Code."* Eleven of the nodes did not cite any earlier nodal work. There is thereby highlighted an implication that these 11 nodes introduce new fundamental information into the area encompassed by the network.

A numeric weighting was assigned each node depending upon the number and type of citation connections to and from the node. The highest nodal value found is for a discovery which Asimov described as the most essential contribution to the historical scheme.

The 1961 Science Citation Index was searched to determine the total count of firstauthor citations to every work listed for each nodal author. Senior nodal authors (the 48 distinguished by Asimov) were cited 5,329 times in the 1961 literature (a mean of 112 citations per author), while junior nodal authors (those not mentioned by Asimov) were cited 1,706 times (a mean of 41.6 citations per author). In the 1961 SCI the average reference author is cited 5.5 times while recent Nobel Prize winners (1962 and 1963) were cited an average of 169 times. More senior than junior nodal authors had citations to works published earlier than the date of the nodal work, and generally the earliest cited work for a senior nodal author predated those for junior nodal authors by a mean of nearly 6 years. This chronological positioning is consistent with the concept that senior nodal authors were more "established" by the time nodal papers were published.

In 71 instances in the 1961 SCI nodal authors cited works by other authors of different nodes. These cases provide evidence for a citation "leapfrogging" effect involving spans of many years. In certain cases leapfrogging reinforced already established historical or citational dependencies between nodes. The frequency of leapfrogging by nodal authors increases sharply among the fourteen most recent nodes --- those representing the coalescence of the new field of molecular biology of the genetic code.

The 1961 SCI revealed that in 58 instances a nodal author cited a work by a co-author. Of the 58 citations, 50 involve citations to the most recent twelve nodes.

The number of citations in the 1961 Science Citation Index to individual nodal articles was compared to those for other articles by the same first author. In a ranked listing half of the cited nodal articles ranked higher than sixth. The nodal work of more than half of the recent (i941-1961) authors ranked as the most heavily cited work for that author. Recent nodal articles also have a higher average absolute count of citations. Therefore not only are nodal authors well cited, but there also exists a strong tendency for their most important works to be cited especially heavily. A special Nodal Citation Index (NCI) was prepared in order to further analyze the bibliographies of nodal papers. In the NCI entries

are repeated for all secondary reference authors, thereby, more easily revealing self-citation pattems and an investigator's possible contribution to one or more other nodes. The NCI also reveals coupling between nodal works which cite the same group of references. This can indicate to what degree any two discoveries are dependent on a mutually shared reference.

The work of twenty-six primary and/or secondary non-nodal investigators found in the NCI was cited by authors of at least three different nodes. Thirteen of these 26 investigators were cited more heavily in the *1961* SCI than the mean for senior nodal authors mentioned by Asimov. Twenty-five of the 26 are cited more heavily than the mean for junior nodal authors. Therefore non-nodal authors cited by at least three different nodes are also well cited in the 1961 literature and are of comparable rank (as measured by citation count) to the nodal authors themselves. Four of the heavily cited references from these 26 non-nodal authors were selected with the aid of additional criteria and investigated for their historical importance. One such reference definitely had the characteristics of a major breakthrough. The others involved innovations in methodology, a difficult matter to evaluate historically. The experiment indicates how even a limited citation index can aid the historian in discovering works not known by him but which should be considered and evaluated. The historian could also profit by considering possible historical implications between nodes connected by citation linkages.

A special Source Index for all the nodal articles arranged by first author was also prepared. This Source Index gives the full authorship of each paper, article title, type of article, the number of authors and works cited by the source paper, the chronological node number, a brief historical description, country of origin of the work, numeric evaluation of citation relationships, organization where the work was done, supporting grants and the complete bibliography.

Fifty-five percent of the nodal research was performed in the United States. The United States Public Health Service and its National Institutes of Health provided grant or fellowship funds supporting 67% of the more recent nodal works (published since 1946).

The average number of authors per nodal paper (2.15) is not significantly different from the average authorship reported for all biomedical papers. The proportion of nodal papers with only one author (16/65) also was undistinguishable from reported averages. Evidence is presented to demonstrate that nodal authors are heavily cited by non-nodal authors and therefore, are in the mainstream of science, yet a certain degree of "cliquishness" among nodal authors is quantitated.

It is concluded that citational patterns provide a valid and valuable means of investigating historical dependencies. Other studies have been suggested for continued research on this subject.

#### **III. INTRODUCTION**

The role of the historian is to describe events and provide perspective on the relationships between events which may seem isolated to the untrained observer.

The reports concerning the assasination of President Kennedy serve well to demonstrate the\difficulty of amassing the "facts" of history even of an event which was observed by countless persons. The data have been analyzed by many experts with great investigative talents. And yet there still remains doubt as to what precisely occurred. It is not surprising therefore, that there are always numerous uncertainties in writing even a fragment of the history of science. The writing of history is subject to much human error in spite of the dedication and relatively rigorous standards held by the professional historian. Unlike legal testimony, motivation and the evolution of ideas are all too often omitted from scientific writings. Tracking down pertinent documents also involves well-publicized difficulties. Historical description must therefore fall far short of an ideal. We can only strive to develop methods that bring us somewhat closer to the truth.

Major achievements in science are relatively easily recognized milestones on the road of progress. However, the minor and less heralded contributions are difficult to identify and even relatively important discoveries may be overlooked in the plethora of data to be evaluated. The historian, in describing the progress of science, is limited by his own experience, memory, and the adequacy of the documentation available. His subjective judgement primarily determines the historical picture of the development of events.

Before World War II the historical perspective of science was relatively easy to gauge. The pace of discovery was slower, scientific fields were less crowded, and the time between basic discovery, evaluation, and application was generally more protracted. Today many new technologies have arisen, and organized research continues to grow at an exponential rate. In sifting the voluminous output of this research, there is an increasing possibility that the historian may eliminate the wheat with the chaff. It becomes ever more difficult to identify potentially important contributions and establish criteria of excellence. The historian's task therefore becomes more complex.

The bibliographies contained in most scientific papers represent a brief history of the subjects they treat and lead to earlier related events. These bibliographies may be usefully reassembled by citation indexing methods in a new chronological orientation -- leading to the later related events. However, analyses based on citation counts must be challenged with the question, "What is the relationship between citation frequency and the historical impact or importance of the work cited?" High citation counts reflect impact but may or may not reflect intrinsic worth. The data obtained from citation analysis are always relative rather than absolute.

In a "citational" approach to historical description one must consider the fact that some scientists consciously or unwittingly ignore earlier work -- at least in their bibliographical

data. Our previous experience using citation indexes for information retrieval as well as the results of the present study indicate this factor is of minor significance, at least when utilizing literature published during the past two or three decades. The refereeing system has undoubtedly helped insure that most pertinent bibliographical data are used in published papers. However, what may be lacking in one paper will be provided in another.

Dr. Isaac Asimov, in his book *The Genetic Code*, has clearly and concisely described the interplay of a century of complex research which led to our present understanding of the DNA genetic code mechanisms for directing protein synthesis. Interspersed in his text are descriptions of milestone discoveries in the history of DNA. Each of these events can be plotted as vertices or nodes in a topological network diagram. Dr. Asimov, writing essentially from memory, did not use the original technical papers or their bibliographies. In his book, he describes some of the specific dependencies of linkages between these nodes or events. Other historial relationships between nodes are implicit in the book or evident through careful interpretation.

In this study, we have investigated in depth the correlations that may exist between Asimov's historical analysis of the key DNA discoveries and a similar analysis derived from citation data covering these same discoveries. The investigation, therefore, is an exploratory comparison of two methods of characterizing history (1) conventional or traditional subjective analysis (2) objective citational or bibliographical analysis.

#### IV. METHODOLOGY

(1) Isaac Asimov's book, *The Genetic Code*, New American Library, New York, 1963, was used as the starting point from which a network schema was constructed which graphically outlines the key discoveries leading to our present understanding of the mechanisms and role of DNA in protein synthesis. (A synopsis of The *Genetic Code* in chapter form is provided for reference in Appendix I). The synopsis has been approved by Dr. Asimov and permission to include the synopsis here was obtained from the publisher, the New American Library.

(2) The key discoveries described by Asimov were plotted as nodes in an historical network schema. Criteria for selection of these nodes from Asimov's text were based on:

(a) A description of discoveries by explicitly named investigators.

(b) A description of discoveries of very obvious importance -- not explicitly named by Asimov, but easily identified due to his provision of other data such as date or place of investigation. For example, Jacob and Monod (Node 35) are described by Asimov as scientists at the Institut Pasteur, Paris, who demonstrated the existence of messenger RNA in bacterial cells in 1961.

Events which were vaguely described were excluded as nodes. Forty nodes were established of which 36 were explicitly named and the balance inferred from Asimov's data. The first node, chronologically speaking, is the work of Braconnot in 1820 and the last that of Nirenberg and Matthaei (1962) -- covering about 140 years.

(3) An extensive literature search using conventional bibliographic tools was completed in order to identify citations for the specific published work described by Asimov for each node. The strictest scholarly criteria were adopted to insure not only that the reference coincided with the node, but also that the reference citation chosen was the paper which most definitely corresponded to the discovery in question. These limitations imposed an important restriction since very often a subsequent work extended the applications of the discovery and established citation connections not to be found in the original paper. (See Appendix II). However, 17 out of 40 nodes in the historical diagram actually represent more than one published paper. Stated another way, several of the nodes on the pure citation network have been coalesced to represent a single node on the Asimov network.

(4) Copies of all pertinent articles were obtained along with translations when these were available. Sixty-five articles were required to cover the 40 nodes explicitly or otherwise described by Asimov. (These are listed in Appendix VI.)

(5) The nodes were plotted chronologically and grouped in broad subject classifications such as nucleic acid chemistry, protein chemistry, genetics, microbiology, or pertinent combinations of these disciplines. Asimov's book was then examined to determine the historical relationships between these 40 nodes. The relationships or connections between the nodes are shown in the first two Network Charts, both of which are colored red. Solid lines on one of the red transparent overlays indicate relationships explicitly specified by Asimov. Broken lines on the other red overlay "represent implied relationships. (These charts are folded inside the back cover).

(6) The bibliography of each node article was examined to determine the citation connections between it and other node papers. If it specifically cited any other nodal article, connecting lines for *direct citations* were established on the Network Charts. The bibliographic examination was extended to include somewhat less direct linkages between the nodes whenever other closely related works by authors of the earlier nodal papers could be found. If a particular node could not be linked to any earlier node by either of these methods, other likely citation pathways were examined, such as connection via an intermediate self-citation, and as a last possibility intermediate connections through any other references cited in the later nodal paper. (Detailed connections are described in Appendix II). In order to facilitate analysis, the network is printed on colored overlays or transparencies which when superimposed emphasize instances of verification by citation analysis of the historical relationships established by Asimov in his book. Thus, the blue overlays show the same 40 nodes described in Asimov's book. The blue solid and dotted lines indicate the existence of reference citations in the nodal papers linking two nodes. For example, Mirsky (39) cites Monod (35). The blue lines are citations which are *coincident* with red lines, that is, indicate where the connectivity of two events explicitly or implicitly described by Asimov are also revealed by a special citation index created for the 65 node papers.

Finally, the yellow overlays show citation connections between nodes which are not disclosed by Asimov. The legend for overlays appears as the last appendix, that immediately preceding the transparencies inside the book cover.

(7) A special citation index based on the 65 papers was created so that pertinent connections between nodal papers could be established. The special Nodal Citation Index (NCI) contains all pertinent data for primary as well as secondary authors. (Appendix III).

(8) In a separate bibliography or Source Index each nodal article is listed and arranged alphabetically by first author. Each item is provided with complete bibliographic data such as full authorship, journal, volume, page, year, type of article, number of authors, and works cited (as well as the complete bibliography itself), chronological node number, title, a brief Asimov description of the node, country of origin, numeric evaluations of citation relationships, organization where the work was done and supporting grants. This bibliography is found in Appendix VI.

(9) Separate listings of the nodal articles arranged by supporting agency, by organizational location of work, and by numeric weighting factor representing the degree of citational relationships were also prepared. (See Appendices V, IV, II)

(10) The 1961 Science Citation Index was searched to determine the total number of citations of every work listed for each nodal author in which he was first author. This information was broken down into self-citations, citations by authors of the same nodes, citations by authors of different nodes and the year of the earliest cited paper. The tabulated material was analyzed to determine if certain authors distinguished by Asimov were subject to citation patterns different from nodal coauthors not mentioned by Asimov and who therefore are implied to be less important. The 1961 Science Citation Index was also examined to reveal any additional citations to nodal authors by other nodal authors. Such data was not incorporated, however, into the overlay sheets. (See p. 7).

(11) The *1961 Science Citation Index* was searched in order to determine the number of citations to each nodal article. On the basis of the 1961 citation counts, the nodal papers were each ranked relative to the other cited works listed for that author. (See p. 15).

(12) The 1961 citation counts for individual papers and authors not mentioned by Asimov (but which were heavily cited in the Nodal Citation Index and therefore might be important) were compared with counts for papers and authors specified by Asimov. The citation relationship between nodal authors within the Nodal Citation Index was also studied. (See p. 23).

#### V. ANALYSIS OF THE CITATIONS TO NODAL AUTHORS FROM THE 1961 SCIENCE CITATION INDEX

What objective support does one find in citation frequency data for the subjective importance which Asimov attributes to the investigators he singled out in the history of DNA? To answer this question, we examined the 1961 *Science Citation Index* and in general found a positive correlation between citation frequency and inclusion in the network. This correlation is similar to that found in another study by us which shows that Nobel prize winners have unusally high citation counts. A large number of the key discoveries named by Asimov were, in fact, made by Nobel prize winners.

The 1961 Science Citation Index was therefore used to analyze citations to authors of nodal articles. There are 89 investigators who served as authors of nodal papers. Asimov, however, mentioned only 48 of these and therefore implies that these men are more important in the scheme of history. For the purposes of this report these men are considered senior authors, while those not mentioned by Asimov (the additional 41 coauthors) are considered junior authors.

It might be expected that, in general, the works of the senior investigators would have been more heavily cited than works by coauthors. In essence, the *1961 Science Citation Index* was used to examine all citations to the works in which any nodal scientist was first author. The following information is tabulated for each author in Table I.

- 1. Total number of 1961 citations.
- 2. Number of citations by non-nodal authors.
- 3. Number self-citations.
- 4. Number of citations by nodal coauthors.
- 5. Number of citations by other nodal authors.
- 6. The publication date of the earliest paper cited.

#### TABLE 1

#### 1961 SCIENCE CITATION INDEX CITATIONS TO NODAL AUTHORS

		Total Number	Citations by	Salf	Citation by	Citation by other	Publication
Node	Author	citations	node authors	Citations	Coauthor	Node author	Paper Cited
40	*Matthaei JH	4	0	1	3	0	1961
40	*Nirenberg MW	28	26	0	1	1	1956
39	Sibatanl A	18	18	0	0	0	1952
39	*Allfrey VE	130	118	11	1	0	1951
39	*Mirsky AE	65	64	0	1	0	1935
39	De Kloet SR	4	3	1	0	0	1960
38	*Novelli GD	32	32	0	0	0	1944
38	Eisenstadt JM	4	3	1	0	0	1959
38	Kameyama T	11	8	0	2	1	1959
37	*Dintzis HM	27	27	0	-	0	1952
36	Bresler A	4	4	0	0	0	1959
36	Diringer R	0	0	0	0	0	-
36	*Hurwitz J	72	65	3	0	4	1952
35	*Jacob F	223	200	20	1	2	1951
35	*Monod J	155	132	2	18	3	1937
34	*Hoagland M	216	204	1	4	7	1954
34	Stephen son ML	10	9	0	0	1	1956
34	Scott JF	22	21	0	Ι	0	1948
34	Hecht LI	85	82	0	2	1	1954
34	Zamecnik PC	101	95	0	3	3	1945
33	*Komberg A	343	336	1	1	5	1942
33	Lehman IR	54	49	0	3	2	1956
33	Simms ES	0	0	0	0	0	-
33	Bessman MJ	49	46	1	1	1	1958
32	Grunberg-Manago M	64	61	0	2	Ι	1953
32	*Ochoa S	165	156	6	0	3	1938
32	Ortiz PJ	5	2	0	2	1	1959
31	*Fraenkel-Conrat H	261	250	5	0	6	1940
31	Williams RC	81	81	0	0	0	1944
30	*Palade GE	449	445	2	1	1	1949
30	Siekevitz P	172	167	0	0	5	1949
30	Porter KR	222	216	6	0	0	1939
29	Michelson AM	99	83	13	3	0	1949

• Senior investigator (mentioned by Asimov)

(continued)

TABLE 1 (cont'd)							
1961	SCIENCE	CITATION	INDEX	CITATIONS	ТО	NODAL	AUTHORS

Node	Author	Total Number of 1961 citations	Citations by non- node authors	Self- Citations	Citation by Coauthor	Citation by other Node author	Publication Year earliest Paper Cited
29	*Todd AR	21	21	0	0	0	1936
28	*DuVigneaud V	145	144	0	0	1	1930
28.	Ressler C	41	36	5	0	0	1953
28	Swan JM	38	34	3	0	1	1952
28	Roberts CW	6	6	0	0	0	1954
28	Katsoyannis PC	41	38	3	0	0	1957
28	Lawler HC	8	6	2	0	0	1953
' 28	Popenoe EA	27	27	0	0	0	1950
27	*Watson JD	111	105	0	0	6.	1950
27	*Crick FHC	118	113	1	0	4	1950
26	*Wilkins MHF	50	50	0	0	0	1951
•26	Randall JT	65	65	0	0	0	1930
26	Stokes	11	I 1	0	0	0	1944
26	Wilson HR	Ι	1	0	0	0	1957
25	*Hershey AD	170	168	0	0	2	1938
25	*Chase M	18	18	0	0	0	1957
24	*Smager F	255	245	10	0	0	1943
24	Tuppy H	61	57	0	4	0	1953
24	Thompson EOP	28	24	3	1	0	1954
23	*Pauling L	630	621	8	0	1	1925
23	*Corey RB	17	17	0	0	0	1936
23	Branson HR	4	4	0	0	0	1950
22-21	*Chargaff E	223	223	0	-	0	1931
20	*Avery OT	56	55	0	0	1	1919
20	*MacLeod	15	15	0	0	0	1940
20	*MeCarty M	90	90	0	0	0	1945
19	Gordon	56	53	3	0	0	1929
19	*Martin AJP	70	70	0	0	0	1940
19	*Synge RLM	39	34	5	0	0	1939
19	Consden R	79	79	0	0	0	1944
18	*Beadle GW	94	94	0	0	0	1931
18	*Tatum EL	45	45	0	0	0	1932

\*Senior investigator (mentioned by Asimov)

(continued)

## TABLE 1 (cont'd)1961 SCIENCE CITATION INDEX CITATIONS TO NODAL AUTHORS

		Total					
		Number	Citations by			Citation	Publication
		of 1961	non-	Self-	Citation by	by other	Year earliest
Node	Author	citations	node authors	Citations	Coauthor	Node author	Paper Cited
17	*Caspersson T	121	115	6	0	0	1924
17	*Schultz J	62	62	0	0	0	1932
16	*Bawden FC	95	89	6	0	0	1933
16	*Pirie NW	67	67	0	0	0	1931
16	Bernal JD	80	80	0	0	0	1924
16	Fanku ehen I	5	5	0	0	0	1933
15,12,9	*Levene PA	147	143	0	3	1	1901
15	Tipson RS	33	27	6	0	0	1939
14	*Stanley WM	18	17	0	-	1	1932
13	*Alloway JL	3	3	0	-	0	1932
12	London ES	7	7	0	0	0	1899
12	Mori T	19	1'4	5	0	0	1949
11	*Griffith F	19	19	0	-	0	1911
10	*Muller HJ	156	121	35	0	0	1914
10	Dippel AL	5	5	0	0	0	1934
9	Jacobs WA	73	73	0	0	0	1915
8,6	*Fischer E	258	256	0	0	2	1878
7	*DeVries H	5	5	0	0	0	1901
6	Piloty 0	13	13	0	0	0	1897
5	*Kossel A	21	20	0	-	1	1888
4	*Flemming W	10	10	0	-	0	1879
3	*Miescher F	6	5	0	-	1	1879
2	*Mendel G	3	2	0	-	1	1865
1	*Braconnot H	Ι	1	0	-	0	1819
		7.025	6 721	175	59	71	
	TOTALS	1.035	6,/31	1/5	58	/1	

• Senior investigator (mentioned by Asimov)

A. Comparison of Senior and Junior Nodal Authors

The average number of authors of nodal papers was 2.15. This value is indistinguishable from the over-all average currently reported in the literature\*. Sixteen papers in thirteen nodes have single authors (37, 22, 21, 14, 13, 11, 8, 7, 5, 4, 3, 2, and 1). Twenty-seven nodes have multiple authors. In seven of those nodes (40, 35, 27, 25, 20, 18, and 17) all the contributing authors are considered senior investigators, i.e., those mentioned by Asimov. This leaves twenty nodes which contain junior coauthors, i.e., those not mentioned by Asimov. For 17 of these 20 nodes the senior investigators are, indeed, more heavily cited than the junior coauthors. The three exceptions are analyzed below:

Node 29 - Michelson is cited more heavily than Todd(99 vs 21). However, the two men were often coauthors. Michelson was usually listed as first author for a series of heavily cited papers (including the nodal reference).

Node 26 - Randall is cited more heavily than Wilkins (65 vs 50). However, if the two men are compared since 1951 (the date of Wilkins' earliest cited papers while Randall's earliest is 1930), Wilkins would be cited more heavily (50 vs 43).

Node 19 - Consden is cited more heavily then Martin (79 vs 70). However, the principal nodal paper (B19) was cited 23 times, and Consden was the first author.

The senior investigators discussed by Asimov, therefore, are generally more heavily cited than their unmentioned coauthors. Another impression seemed evident regarding the unmentioned coauthors; most were cited more heavily during years following the publication of the nodal articles to which they had contributed.

As a base line for the discussion which follows it should be noted that the average reference author in the 1961 SCI was cited 5.5 times while the 13 Nobel prize winners in physics, chemistry, and medicine for 1962 and 1963 were cited an average of 169 times.

B. Breakdown of the Total Count by Type of Citation

Of the 7,035 citations in the 1961 Science Citation Index to all nodal authors:

- 1. 5,329 citations were to 48 investigators discussed by Asimov -- a mean of 112.0 citations per author.
- 2. 1,706 citations were to 41 co-investigators -- a mean of 41.6 citations per author.
- 3. There are only 175 self-citations by 30 of the 89 nodal authors in the entire 1961 SCI. (First author citing first author is a self-citation here) It should be noted that the chronological span for this history is 140 years, therefore, only the more recent nodal authors could possibly be involved in self-citations in 1961. If only authors involved in nodal discoveries since 1935 (Node 14) are considered, the statistic reads 135 self-citations by 28 of the 74 authors. A notable exception in the earlier group is Herman Muller whose work at age 71 spans half a century. Therefore, an analysis of the current self-citation practice and the date of the earliest paper cited provide an obvious measure of the extent of an author's

\*Clarke, B.L., Science 143:822 (1964) - (See Reference 7, p. ii)

involvement in the history of his field.

- 4. In 1961 there were 58 instances in which a nodal author cited a work in which one of his nodal coauthors was first author. These citations most frequently involve coauthors of nodes 29 to 40 (or from 1955 to 1961) since 50 of the 58 citations are for this period.
- 5. There are 71 instances in the 1961 SCI in which nodal authors have also cited various works in which the authors of other nodal works were first authors. This may enable us to provide a new method of demonstrating historical correlations through retrospective analysis.

C. Retrospect: The 1961 Citation of a Nodal Author by the Author of a Different Node

It is possible that two nodal works have no parallel relation to each other until both their contributions were eventually utilized by future investigators. For instance, it is difficult to historically relate nodal work by Muller (10) 1926 and Levene (12) 1929 because of the dissimilarity of their work at a period which had no indication for establishing relevance. It can be assumed also that no citation linkage (or at best a rather tenuous difficult-to-establish citation linkage) exists between the two nodes, that is, node 12 to node 10. Yet in 1961 Muller cites a work by Levene. It must be assumed that a relevance has now been established by Muller, albeit in retrospect.

This example and others may establish a connection where none were demonstrated by Asimov or by citation indexing of the nodal papers. It is important to reiterate that this study could not determine whether in fact citation linkages exist that might have been found with a more comprehensive citation index accumulated across many source years. Other instances however, actually coincide with connecting citation lines shown on the historical network chart. The original chronological relationship is reversed in 31 of the 71 citations which are outlined in detail below.

1. Early nodal authors citing a general work by recent nodal authors in the 1961 Science Citation Index (underlining of the node number indicates agreement with citation connecting lines between two nodes on the historical network chart):

Hoagland	(34)	cites	Jacob (35)
Ochoa	(32)	"	Hurwitz (36) 2x, Hecht (34), Kornberg (33) 2x
Todd	(29)	"	Kornberg (33), Ochoa (32), Watson (27)
Crick	(27)	"	Nirenberg (40), Jacob (35)
Sanger	(24)	"	Fraenkel-Conrat (31) 2x, Du Vigneaud (28), Swan (28)
Tuppy	(24)	"	Fraenkel-Conrat (31) 2x
Synge	(19)		Stephenson (34)
Stanley	(14)	"	Hoagland (34), Watson (27), Crick (27)
Muller	(10)	"	Hoagland (34), Lehman (33), Ochoa (32), Fraenkel-
			Conrat (31), Watson (27) 2x, Crick (27), Hershey
			(25), Avery (20), Levene (15)

The chronological relationship is unchanged in 40 of the 71 citations listed below.

2.	2. Recent nodal authors citing a general work of early nodal authors in the 1961					
	Science Citation I	<i>ndex</i> (u	nderlinin	g of the node number indicates agreement with		
	citation connectin	g lines	between	two nodes on the historical network chart):		
	Nirenberg	(40)	cities	Hoagland (34), Siekevitz (30), Hershey (25)		
	Matthaei	(40)	"	Kameyama (38), Hurwitz (36), Hoagland (34),		
				Siekevitz (30) 2x		
	Allfrey	(39)	"	Monod (35), Hoagland (34), Zamecnik (34) 2x,		
				Kornberg (33), Palade (30)		
	DeKloet	(39)	"	Hoagland (34) 2x, Siekevitz (30)		
	Novelli (38) "			Hurwitz (36), Monod (35) 2x, Zamecnik (34),		
				Siekevitz (30)		
	Hurwitz	(36)	"	Lehman (33), Bessman (33), Grunberg-Manago		
				(32), Ochoa (32), Ortiz (32), Watson (27)		
	Jacob	(35)	"	Kornberg (33), Crick (27)		
	Monod	(35)	"	Crick (27), Pauling (23)		
	Ochoa	(32)	"	Fraenkel-Conrat (31)		
	Fraenkel-Conrat	(31)	"	Stanley (14)		
	Todd	(29)	"	Watson (27)		
	Synge	(19)	"	Fischer (8,6)		
	Tipson	(15)	"	Fischer (8,6)		
	Muller	(10)	"	Kossel (5), Miescher (3), Mendel (2)		

Analysis reveals 29 instances in which citation connections between two nodal authors (expressed in the 1961 SCI) agree with citation connections formed between the same nodal authors on the historical network chart. Forty-two additional citational connections not found on the historical network chart are also demonstrated.

It is important to note here that indirect citation linkages can undoubtedly be demonstrated between nodal papers which, in our blue and yellow transparencies, are not connected. The use of larger citation index files extending over many source years would probably disclose non-nodal "stepping stones" between most of these "unconnected" nodes. D. Citation Leapfrogging Effect

The chronological relationships in parts 1 and 2 above evidence a citation leapfrogging effect across a span of many years. For example, analysis of nodal papers shows that Hurwitz (Node 36) 1960 cites Ochoa (Node 32) 1955-56; however, in 1961 Ochoa cites Hurwitz (and Hurwitz again cites" Ochoa). Other citations between both men may exist and would be discovered by a comprehensive citation analysis of all their works.

Analysis of the frequency with which certain nodal authors are cited in 1961 by other nodal authors is an indication of their involvement in this leapfrogging phenomenon. This frequency (number of times involved) is plotted against the nodal numbers (i to 40) in the

following histogram. There is a sharp increase of involvement in citation leapfrogging that begins with Watson and Crick (Node 27) whose work, published in 1953, advance an important theory of nucleic acid structure. This increase in frequency coincides with the event which one might intuitively call the coalescence of a new subfield, namely, the molecular biology of the genetic code. This method of recent (1961)citation patterns between nodal authors also appears to pinpoint that event which Asimov as an historian describes as the "... model which finally made sense of all the data that had been painstakingly collected on purine and pyrimidine ratios, and which was destined to make immediate sense of the problem of replication ... "

E. Chronological Position: An Analysis of the Earliest Cited Work by a Nodal Author The date of the earliest eited work by a nodal author also provides chronological perspective to the nodal paper. Of the 48 senior nodal authors distinguished by Asimov, only four (Chase, De Vries, Miescher, and Kossel) did not have cited works in the 1961 Science Citation Index which were earlier than their nodal dates. Eleven of the 41 secondary nodal coauthors were not cited for papers earlier than their nodal dates.

For the 44 senior nodal authors who had earlier works cited the average difference between earliest paper and the nodal paper is 12.4 years, and the median is 11 years.

Similarly, for the corresponding group of 30 secondary nodal authors the average difference is 6.8 years and the median is 5 years. Therefore, senior nodal authors appear to be more "established" than their coauthors by the time nodal papers are published.

From the above results it seems evident that citation indexing objectively supports, with quantitative data, the subjective emphasis that an historian has placed on the contributions of the distinguished authors. Futhermore, many of those involved in past discoveries and who remain active continue to reinforce past nodal author interdependencie in the bibliographies of their most recent works.



#### VI. ANALYSISOF THE CITATIONS TO NODAL ARTICLES FROM THE **1961** SCIENCE CITATION INDEX

#### A. Selection of the Nodal Article

Sixty-five articles are associated with the forty nodes of this study. These were identified after an extensive literature search of the subject and author indexes in *Chemical Abstracts, Current List of Medical Literature, Cumulative Index Medicus,* etc..

The initial search revealed many candidates for certain nodes. Each candidate paper was critically reviewed in order that the subject content would agree as closely as possible to Asimov's description. Generally, the more difficult choices occurred in papers which were published in the last fifteen years (since 1945) of the period described in Asimov's history. There are two reasons for this difficulty: (1) Lately, communication of a significant discovery is frequently presented in several sources within a very brief period, (2) certain significant contributions involve numerous sequential stages in their evolution and recently the trend seems to be to publish after each stage is completed. This makes it difficult to determine exactly in which paper the concept is originally established or proven. For example the nodal paper for Todd (Node 29) is part *32* in a series.

As a consequence of these difficulties there are certain prerequisites for attempting this type of network study. These include considerable experience and competence in using and searching the literature, and a post-graduate level of training (or its equivalent) in the subjects reviewed by the history. Otherwise, the choice of nodal papers could be poor, introduce serious distortions, and lead to false conclusions.

The limitations imposed by the search-selection are controls required to test the citation network under rigid conditions. For instance, the Watson and Crick discovery of the molecular configuration of DNA consisted of two articles published in the 1953 volumes of *Nature*. The bibliographies contained in these papers were extremely brief and seemingly of little value in demonstrating citation dependency on earlier work. Within the year, Watson and Rich published a brief paper (*Proc. Nat. Acad. Sci. U.S.* 40:759, 1954) on the same subject which, unlike the two previews papers, directly cited nodal articles by Avery et al (20), Hershey and Chase (25), Wilkins (26), and Chargaff (22). There were other papers which also demonstrated many more connections to nodal articles than did the earliest paper which fully described the discovery. The present report, therefore, does not attempt to demonstrate the blunt force of numerous citations from "convenient" papers; it tries rather to analyze the citation linkages which play a more meaningful role in the historical evolution of the subject.

#### B. Ranking of Citation Counts to the Nodal Article

In the Table 2 the sixty-five nodal articles'are listed by their first author. The *1561 Science Citation Index* was consulted to determine the number of citations to each

paper• This figure was compared to the number of citations for other individual papers by the same author in which he was first author, and a relative ranking established.

#### TABLE 2

#### Ranking of Nodal Articles Relative to Other Cited Works by the Same First Author Based on Citation Counts Found In 1961 (or 1964) Science Citation Index

Nodal Articl	es	1961 SCI	Ranking by
(Arranged		Number of	Citation
Chronologicall	y)	Citations I	Count 2
1961-2			
Matthaei	A40	30*	1
Nirenberg	B40	112"	1 (1)
Nirenberg	C40	10"	2 (1)
Sibatani	A39	40*	1
Novelli	A38	1	> 5
Eisenstadt	B38	7*	1
Kameyama	C38	4*	> 1
Dintzis	37	iO	1
Hurwitz	36	23	1
Jacob	35	24	1
Hoagland	A34	27	3
Hoagland	B34	57	1 } (1)
Kornberg	A33	1	> 5
Kornberg	B33	2	> 5 } (>5)
Kornberg	C33	6	> 5
Grunberg-Manago	A32	6	4
Grunberg-Manago	B32	13	2 } (2)
Ochoa	C32	2	> 5
Fraenkel-Conrat	A31	9	3
Fraenkel-Conrat	B31	11	2 } (2)
Fraenkel-Conrat	C31	6	> 5
Palade	A30	14	> 5
Palade	B30	43	3 } (2)
Michelson	29	3	> 5

1 Asterisk indicates number of citations in the 1964 SCI.

2 Number in parentheses is rank if citations to papers by the same first author are totaled and treated as one paper.

	Nodal Ar	ticles	1961 SCI	Ranking by
	(Arranged		Number of	Citation
	Chronologi	cally)	Citations	Count 2
	DuVigneaud	A28	5	> 5 (2)
	DuVigneaud	B28	8	$3^{(2)}$
	Watson	A27	44	1
	Watson	B27	27	$2^{(1)}$
	Wilkins	A26	3	5
	Wilkins	B26	5	2 } (2)
	Hershey	25	31	1
	Sanger	A24	15	4
	Sanger	B24	17	3
	Sanger	C24	24	$2^{(2)}$
	Sanger	D24	11	> 5
	Pauling	A23	5	> 5
	Pauling	B23	25	4 } (1)
1951	Pauling	C23	5	> 5
1751	Chargaff	22	1	> 5
	Chargaff	21	0	> 5 } (>5)
	Avery	20	33	1
	Gordon	A19	1	> 5
	Consden	B19	23	1
1941	Beadle	18	7	3
	Caspersson	A17	1	> 5
	Caspersson	B17	1	$> 5^{\{(>5)\}}$
	Bawden	A16	0	> 5
	Bawden	B16	3	5 } (5)
	Levene	15	0	> 5
	Stanley	14	0	> 5
	Alloway	13	2	i
	Levene	A12	2	> 5
	Levene	B12	0	> 5 } (>5)
	Griffith	11	10	1

2 Number in parentheses is rank if citations to papers by the same first author are totaled and treated as one paper.

Nodal Arti	icles	1961 SCI	Ranking by
(Arrange	d	Number of	Citation
Chronologic	ally)	Citations	Count 2
Muller	10	0	> 5
Levene	A 9	0	$> 5_{1(>5)}$
Levene	B 9	1	$> 5^{1(>3)}$
Fischer	8	0	> 5
Devries	7	0	> 5
Fischer	6	1	> 5
Kossel	5	0	> 5
Flemming	4	1	2
Miescher	3	1	1
Mendel	2	11	1
Braconnot	1	0	> 5
	TOTAL	. 674	

TABLE3

Chronological Summary of Table 2

Nodal Articles Pub-	Average Number of	
lished in the	Citations per Article	
Period	(only from 1961 SCI)	Range
1951-1961	15.1	0-57
1930-1950	5.5	0-33
18,19-1929	1.1	0-11

2 Number in parentheses is rank if citations to papers by the same first author are totaled and treated as one paper.

#### TABLE 4 Breakdown of all 65 Nodal Articles 1819-1962

Ranking of Nodal articles	
relative to other works by	No. of occurrences
same first author	of each ranking
1	17
2	7
3	6
4	3
5	2
> 5	32

#### TABLE 5

Table 5 below demonstrates that there are more instances in recent years in which the Nodal article is the most heavily cited work among those for which the Nodal author was first author.

Breakdown of the Most Recent 44 Nodal Articles 1941-1962

Ranking of Nodal articles	
relative to other works by	No. of occurrences
same first author	of each ranking
1	13
2	6
3	6
4	3
5	1
> 5	15

The above rankings treat each nodal article separately. However, if name repetitions are excluded and we use the parenthetical values from Table 2, there are only 41 individuals who function as first author within the network. We total the citations for each of the 41 individuals and compare each total-to the number of citations given other references by this author. For instance, DuVigneaud's nodal article (A28) was cited five times (Rank 5) in the 1961 *Science Citation Index*. DuVigneaud (B28) was cited eight times (Rank 3). The total of 13 citations (pooling DuVigneaud's nodal articles) would give a new composite ranking of 2. In this sense, both nodal articles are treated as one, and the citation count compared to the number of citations given all other references by the author. This treatment is valid to the extent that later authors will cite only *one* reference out of *several* that have essentially the same context. Furthermore, some of the nodal articles are brief reports of correspondence and herald the subsequent nodal paper containing more substance. For example, articles A16, A19, A23, and A38 are brief preliminary letters which all rank >5.

#### TABLE 6A

#### Citation Ranking of Pooled Nodal Papers for 41 Nodal First Authors 1819-1961

Ranking of Nodal articles		
relative to other works by		No. of occurrences
same first author		of each ranking
1		18
2		7
3		1
4		0
5		1
>5		14
	TABLE 6B	
	1941-1961	
Ranking of Nodal articles		
relative to other works by		No. of occurrences
same first author		of each ranking
i		14
2		6
3		1
4		0
5		0
>5		6
	TABLE 6C	
	1819-1941	
Ranking of Nodal articles		
relative to other works by		No. of occurrences
same first author		of each ranking
1		4
2		1
3		0
4		0
5		1
>5		8

Another adjustment is possible; papers ranked >5 can be excluded if a different first author has written another paper (in the same node) which ranks 1-5. The 1941-61 group would thereupon drop three authors whose papers ranked >5 (Table 6D).

Ranking of Nodal articles	
relative to other works by	No. of occurrences
same first author	of each ranking
1	14
2	6
3	1
4	0
5	0
>5	3

#### TABLE 6D

#### Adjusted 1941-1961

The three remaining authors whose nodal works ranked >5 in Table 6D are Chargaff, Kornberg, and Michelson who are among the more heavily cited authors in nucleic acid chemistry. Their more current work continues to generate such interest that they are cited more often than references six to ten years old. Also Chargaff and Michelson are editors and authors of recent text references on nucleic acid which arc cited very heavily and contain, in essence, a review of their nodal discoveries. The ranks of many nodal articles would be improved if their citation counts were compared to other references occurring only within the period three years before or after the nodal date. For instance, the 1953 Sanger nodal article (C24) receiving 24 citations, ranks second to a 1945 non-nodal reference by Sanger with 84 citations. However, the top ranking article antedates the nodal discovery by about eight years. Therefore, if workable limits (on the basis of highest number of citations in the *Science Citation Index*) can be imposed on dates, there is increased probability of selecting the most significant article by a given author on a given subject.

It is obvious that recent nodal articles in the network (1941-1961) receive a better relative rank than older articles (Table 5) and, also, the more recent references have a higher average absolute count of citations (Table 3). Over fifty per cent of all nodal articles ranked between one and five (Table 4). Table 6B demonstrates that the nodal work of over fifty per cent of the recent (1941-1961) authors ranked as *the* most heavily cited work by that author.

In evaluating the data in Tables 6A to 6D one must keep in mind that there is generally a higher percentage of citations in the SCI for any single year to papers published during the past few years. This is, in part, due to the fact that there is more recent literature that can be cited. Statistical data on the chronological distribution of reference citations can be found in the *Introductions* to the *1961* and *1964 Science Citation Index*. The use of citation data from any single source year is inevitably biased by the tendency to cite more recent papers.

#### VII. DISCUSSION OF THE CITATION INDEX PREPARED FROM THE SIXTY-FIVE NODAL PAPERS (NCI)

The complete Nodal Citation Index (NCI) is found in Appendix III. This NCI includes entries for every reference work cited in any of the 65 nodal papers. Following each of the numerous cited references there is a brief identifying description for each citing nodal paper. A *complete* description of every nodal document is provided in the Source Index of Nodal Papers (see Appendix VI).

A. The Nodal Citation Index (NCI) as a Method of Historical Investigation

In contrast to the 1961 Science Citation Index which draws exclusively upon source articles published in a single year (1961), the NCI is derived from articles published in various years during the past century. Thus, the NCI is not chronologically restricted. However, the NCI is a derivative of Asimov's text and, therefore, reflects his opinion as to which are the milestone achievements. It was possible however that the papers covered by Asimov cited other important investigators which he does not cite. To investigate this possibility, we determined if non-nodal papers and non-nodal authors heavily cited in the NCI were also heavily cited in the 1961 Science Citation Index. The number and pattern of 1961 citations to distinguished nodal authors and articles have been established in the preceding sections. It was of interest to determine if these heavily cited non-nodal authors or papers had comparable patterns.

If so, then certain heavily cited *authors* and *articles* should perhaps have been included by Asimov in his book.

1. Selection of Articles Cited by at Least Three Separate Nodes

The only non-nodal *article* in the NCI that was cited by at least three distinct authors of three separate nodes was:

Siekevitz P, "Uptake of Radioactive Alanine *in vitro* into Proteins of Rat Liver Fractions," *J. Biol. Chem.* 195,549 (1952). It was cited by Kameyama (38), Nirenberg 2x (40), Palade (39), and Matthaei (40).

Siekevitz also appears as a junior nodal coauthor (not mentioned by Asimov) with Palade (Node 30). His general works received 172 first author citations in the 1961 Science Citation Index which is above the mean of 112 citations for senior nodal authors. The 1952 Siekevitz article received 28 citations in the 1961 SCI and was his most heavily cited paper, as is typical of nodal papers. Siekevitz's method for dealing with the uptake of radioactive alanine in liver microsome fraction was used (and referred to in three nodal articles) as a step in the experimental procedure--the washing and counting of radioactive protein precipitates. The method described by Siekevitz was obviously useful but from an historical point of view it can be questioned whether this discovery constitutes a major discovery.

#### TABLE 7

NCI and 1961 SCI Citation Analyses for Non-Nodal

#### Authors Cited by at Least Three Different Nodes

Nodal Citation Index

#### 1961 Science Citation Index

Non Nodel Authors	No. of Nodas	Num Times (	ber of			Number of NCI	X if any of These Entries	Number of	No. of 1961	Publication Voor
Cited by Three or	Penresented	First	Second		No. of Entries	First Author Entries	Citation Pank*	Citations As	Citations by	of Earliest
More Different Nodes	At Least Once	Author	Author	Total	As First Author	Appearing in SCI	1 or 2	First Author	Nodal Author	Paper Cited
Astbury WT	4	6	0	6	6	3		85	0	1926
Benzer S	3	3	0	3	3	3		135	11	1948
Berg P	3	3	Ι	4	3	1		99	2	1953
Brachet J	4	7	0	7	7	4		347	1	1931
Carter CE	3	6	0	6	5	4	X(25)	48	0	1945
Cohen SS	3	4	1	5	4	2		186	2	1940
Colowick SP	3	1	3	4	1	1	X(16)	199	0	1942
Davidson JN	4	0	I1	11	0	-	Not applicable	101	0	1939
Gros F	4	2	6	8	2	2		136	0	1946
Hammarsten E	3	7	Ι	8	5	4	X (7)	28	0	1924
Heppel LA	3	6	4	10	4	4		119	0	1939
Hulbert RB	3	3	2	5	3	3	X(65)	86	1	1944
Kirby KS	4	5	0	5	2	2	X(43) X(28)	118	6	1955
Lipmann F	3	0	4	4	0	-	Not applicable	189	3	1930
Magasanlk B	4	3	5	8	2	2		100	2	1948
Markham R	3	7	3	10	5	5	X(50)	247	1	1942
Meselson M	3	0	4	4	0	-	Not applicable	98	3	1957
Potter VR	4	2	9	11	2	1		166	0	1941
Rich A	3	5	2	7	4	2	X (9)	114	4	1951
Roberts RB	3	0	5	5	0	-	Not applicable	128	2	1949
Schmitz H	3	2	1	3	1	1	X(11)	49	0	1920
Sevag MG	3	4	1	5	3	2	X(20) X(18)	53	0	1934
Spiegelman S	4	1	6	7	Ι	Ι	X (4)	78	3	1942
Volkin E	5	4	2	6	4	3	X(18)	90	5	1951
Weiss SB	4	3	1	4	3	3	X(27)	108	7	1955
Zamenhof S	4	9	3	12	8	2		151	0	1940

\*Number enclosed in parentheses indicates number of citations

- Selection of Non-Nodal Authors Cited by at Least Three Separate Nodes. Aside from a specific paper like the Siekevitz article the *general work* of 26 non-nodal investigators was cited frequently--that is, by at least three separate nodes. (See Table 7). Four of the 26 well-cited non-nodal authors appear only as secondary reference authors, five only as primary authors, and in 17 instances the position is mixed.
  - a. Comparisons to Nodal Authors

These 26 investigators were studied by examining the 1961 Science Citation Index. Their citation counts were compared with citation counts for nodal authors. Thirteen of the twenty-six investigators were cited more heavily than the mean (112 citations) value for 48 senior (first) nodal authors named by Asimov. Twenty-five of the twenty-six were cited more heavily than the mean (41.6 citations) for 41 junior nodal coauthors. Thus, the non-nodal authors cited by at least three different nodes are also well cited in the 1961 literature and are of comparable rank (as measured by citation count) to the nodal authors themselves.

Excluding self-citations, it is important to note that only 19 of the 48 senior *nodal* investigators in the NCI (Table 8 below) are cited by authors of three or more other nodes. Therefore, this characteristic does not have absolute importance even among nodal references. Our subjective impression from Table 8 is that those nodal authors who are heavily cited by nodal scientists tend also to be the most generally renowned researchers. Note that 39 of the 48 senior nodal authors are cited at least once by another nodal author.

We note at this point that although self-citations should be eliminated from counts used in evaluating the impact of a scientists' work on others, the self-citation linkage to later work by the same author is completely legitimate and is as valid as any other citation in establishing conceptual continuity of research.

#### TABLE 8

#### The Number of Different Nodes Involved at Least Once in the Citation of a Senior Nodal Author

	No Nodes					
	Except for				4 or Mo	re
No Nodes	Self-Citations	1 Node	2 Nodes	3 Nodes	Nodes	
Beadle	Dintzis	Alloway	Caspersson*	Corey	Allfrey	(4)
Bracconot	Du Vigneaud	Bawden	Fraenkel-Conrat	Crick	Avery	(5)
De Vries	Flemming	Chase	Griffith	Fischer	Chargaff	(6)
Mendel	Kossel	Hershey	Jacob	Hoagland	Mirsky	(4)
	Sanger	Muller	Kornberg	Hurwitz	Novelli	(4)
		Palade	Matthaei	Levene	Ochoa	(5)
		Panling	McCarty	MacLeod	Stanley	(4)
		Tatum	Nirenberg	Martin	Watson	(5)
		Wilkins	Schultz	Miescher		
			Synge	Monod		
			Todd	Pirie		

\*(Example: Some author of each of two different nodes cited Caspersson at least once..)

The 26 non-nodal authors in Table 7 were studied further to determine whether any should have been mentioned by Asimov and thereby become nodal authors. Some of the 26 are prominent in the field of nucleic acids. Chargaff, for example, in his nodal article (Node 22) considers the work of Brachet and Hammarsten as important as that of Avery (Node 20) and Caspersson (Node 17). Chargaff in his nodal paper (22) states that Brachet and Hammarsten were "responsible for the enormous revival in interest for the chemical and biological properties of nucleic."

b. Selection of Potential Nodal Articles

In our analysis of the 1961 SCI Citations to Nodal articles, it was shown that nearly 70% of the more recent (1941-1961) nodal articles were the most (or second most) heavily cited articles for the first author in *the Science Citation Index*. From Table 7, one finds four authors who (1) are cited in the 1961 SCI more than 112 times *and* (2) have published a paper which is cited in a nodal paper *and* (3) is the author's most or second most heavily cited article in the *1961 Science Citation Index*. On this basis, the following four specific papers by Colowick, Kirby, Markham and Rich would have qualified as nodal articles in the historical network. Therefore, these four references were studies in further detail:

 Colowick S.P. & Kalckar H.M., "The Role of Myokinase in Transphosphorylations. 1. The Enzymatic Phosphorylation of Hexoses by Adenyl Pyrophosphate," *I. Biol. Chem.* 148,117 (1943).

*Abstract:* In the Embden-Myerhof pathway of glucose (hexose) metabolism hexokinase catalyzes the following reaction:

Adenosine triphosphate (ATP) + hexose hexokinase phosphate. hexokinase adenosine (ADP) + hexose hexokinase

If adenosine dlphosphate ADP were substituted as the phosphate donor the above reaction would not go to completion. However, if myokinase were added to either system the yield would include adenylic acid (AMP) and hexosemonophosphate, since myokinase, with hexokinase, will catalyze the reaction:

myoklnase

ADP + hexose hexokinase AMP + hexosemonophosphate

On the basis of this phenomenon the authors further investigated the action of myokinase on adenine nucleotides. They describe a reaction called "phosphate dismutation" in which myokinase catalyzes the transfer of a labile phosphate from one molecule to another. 2 ADP-----> 1 ATP+ 1 AMP

Sixty per cent of the ADP is converted into ATP and AMP in this simple equilibrium. CITATION NOTES

Colowick's work was cited by Ochoa (34), Kornberg (33), and Kameyama (38). This specific paper by Colowick was cited by Kornberg (33) and cites one work by a nodal author, Levene (15).

 Kirby K.S., "A New Method for the Isolation of Ribonucleic Acids from Mammalian Tissues, "Biochem. J. 64,405 (1956).

*Abstract:* Ribonucleic acid (RNA) was separated from various tissues by a method which permitted extraction with phenol and water at room temperature at pH 6.0-7.5. Pancreatic ribonuclease was inactivated by the same phenol treatment. Most important, however, was that deoxyribonucleic acid remained completely insoluble under the conditions used. This allowed that nuclei did not have to be separated from cell preparations. Also, RNA could be extracted from the DNA - Laden Nucleus.

#### CITATION NOTES

Kirby's work was cited in the Nodal Citation Index by Hoagland (34), Hurwitz (36), Eisenstadt (38), and Sibatani (39). This specific paper by Kirby was cited twice by Hoagland and was apparently essential for his method. Kirby's paper does not cite any nodal authors.

3..Markham R., Smith J.D., "The Structure of Ribonucleic Acids."

1. Cyclic Nucleotides Produced by Ribonuclease and Alkaline Hydrolysis, *Biochem. J.* 52,552 (1952).

*Abstract:* The authors state that ribonuclease degradation of RNA polynucleotide discriminates between purine and pyrimidine nucleotides While alkaline hydrolysis does not. Ribonuclease can be used with easily controlled reactions to provide sufficient nucleotides for study and determination of their structure, and also their sequence in the chain. Electrophorectic methods are discussed.

#### CITATION NOTES

Markham's work was cited by Michelson (29), Ochoa (32), and Sibatani (39). This specific paper was cited by Ochoa (32). The paper cites works by nodal authors Todd, Levene, and Komberg.

4. Rich A., Davies D.R., "A New Two Stranded Helical Structure: Polyadenylic Acid and Polyuridylic Acid," J. Am. Chem. Soc. 78,3548 (1956). [Letter to Editor]. Abstract: Strands of synthetic polyuridylic acid when mixed ~qth strands of synthetic polyadenylic acid formed a helical structure (studied by X-Ray diffraction) containing two strands, one of each type, of nucleic acid. This for the first time shows that RNA can arrange itself in a structure similar to DNA which could account for RNA replication in plant and smaller animal viruses (which contain no DNA).

#### CITATION NOTES

Rich's work was cited by Ochoa (32), Hoagland (24), and Nirenberg (40); the specific Rich paper cites three node papers: Watson and Crick (27), Wilkins (26), and Ochoa (32).

c. Evaluation of Potential Nodes

The papers by Kirby and Markham are cited for their method. The method by Kirby as described by Hoagland represents a very significant improvement since by Kirby's method RNA could easily be separated from DNA even in the nucleus. The methods described by Markham, at least as indicated by those citing him may not be considered a 'major' contribution. The Colowick paper describes the original instance of *in vitro* enzymatic phosphorylation of a nucleotide (ADP -----> ATP), and is cited for this reason. The paper by Rich has the characteristics of a major breakthrough since it describes a

phenomenon which might explain replication of RNA virus -- an enigma which challenged the entire DNA theory. The paper by Rich would seem to qualify for inclusion as a node. The papers by Kirby and Colowick are important but are not as clearly essential to the network. The paper by Markham appears even less essential to this particular network, though, its general value might be considered of greater importance in a history of biochemistry. However, it is not easy to evaluate the historical contribution of methodological discoveries. Methodology, of course, provides the tools for discovery. Carter, Magasanik, Sevag, Volkin and others of the 26 heavily cited non-nodal authors are cited on the basis of their innovations in methodology. Consequently, it appears that it may be useful to construct historical networks of science in such a fashion as to easily characterize the method papers. Perhaps insufficient importance has heretofore been attributed to methodology in writing the history of science. Certainly, in the history of technology, methodology should prove to be an even more important factor.

3. Coupling of Nodal Articles as Demonstrated in the NCI.

As a side excursion into bibliographic coupling we examined one example where non-nodal articles are cited by the same two nodes (32 and 33). Asimov has stated that Ochoa (32) and Kornberg (33) did related work, and indeed they cite each other. Both shared the 1959 Nobel Prize in .Medicine and Physiology.

In the Nodal Citation Index, 19 authors were cited by Node 32 alone, 14 authors by Nodes 32 and 33, and 37 authors by Node 33 alone. We point out the possibility of extending the coupling study to a full evaluation of all the combinations of two and three nodal papers and comparing the quantitative results with subjective and historical impressions of "relatedness" of papers.

4. Intermediate References Used in Indirect Citation Connections

In all cases of indirect citation whether strong or weak (broken lines on blue or yellow overlays) non-nodal journal references were used as intermediate papers in establishing indirect citation connections between the indicated pairs of nodes on the historical network chart. As it turned out, none of the intermediate references we examined could be used as intermediates between any nodes other than the one pair under consideration.

B. Historical Network Chart

Examination of the overlays demonstrates the number of various types of connections between nodes which have been described in the text. (Consult legend on page 74.)

Asimov's Historical Connections	Specified	29
	Implied	14
	TOTAL	43
Coincident Citation Connections	Direct	15
	Strong Indirect	7
	Weak Indirect	6
	TOTAL	28
Non-Coincident Citation Connections	Direct	10
	Strong Indirect	16
	Weak Indirect	5
	TOTAL	31

Thus, there is citation coincidence found in 28/43 of Asimov's historical connections or a coincidence of 65 per cent. These are represented by blue lines. There are 31 additional non-coincident nodal citation connections whose meanings range from perfunctory acknowledgment of an earlier work to a strong dependency on the earlier work not described by Asimov. We note that there are 29 historical connections specified by Asimov and a similar value of 25 (15 + 10) instances in which one node directly cites another.

It might be interesting to examine an historical narrative based on a description of the direct citation linkages and compare that essay with Asimov's original version. C. Lack of Early Citation Dependency and Scientific Originality

The Historical Network Chart also includes eleven papers which might appear to involve no citation dependency on any earlier nodal papers. Only three of the eleven are assigned specific early connections by Asimov; and only one has an earlier implied historical connection. Therefore, seven of the eleven papers are confirmed as starting points which, within this network, have neither a citation nor historical dependency on earlier works. Each of these eleven papers proved to involve highly original work. Node

#### Discovery Reported

- (1) Braconnot isolates the first amino acids.
- (2) Mendel demonstrates the laws of inheritance.
- (3) Miescher isolates nucleic acid.
- (6) Fischer and Piloty determine the structure of ribose, later found to be the carbohydrate fragment of nucleic acid.
- (7) De Vries expresses the concept of natural mutation.
- (10) Muller produces mutations with x-rays.
- (11) Griffith demonstrates bacterial transformation.
- (14) Stanley crystallizes virus.
- (19) Martin and Synge develop the powerful analytical method of paper chromatography for application in protein chemistry.
- (23) Pauling and Corey demonstrate the helical structure of protein.
- (26) Wilkins analyzes nucleic acid by X-ray diffraction.

These works (Nodes 1,2,3,6,7,10,11,14,19,23 and 26) appear to represent key breakthroughs which either present new fundamental information in the evolving field or describe new applications of information from other disciplines. The network that can be derived from an account by an historian highlights those events which that historian considers as fundamental. The lack of backreaching historical reference made evident by the drawing of an historical network facilitates a reevaluation of the historian's assumption of fundamentality. In addition to a subjective reevaluation, one may try to confirm or contradict the assumptions of fundamentality by looking for citation linkages from these "fundamental papers" back to other nodal works. Of course, the earlier a work appears in the chronological network, the less likely it is that one will find citations back to other nodal papers.

#### VIII. CONCLUSIONS

(1) The senior investigators responsible for the nodal papers examined in this study are, on the average, cited in the 1961 *Science Citation Index* with a frequency (112 citations/author) that compares with those for recent winners of the Nobel prizes in science (169 citations/author). Both frequencies are well above the average value (5.51 citation,/ author) encountered in the 1961 Science Citation Index. The frequency of 112 citations/' author is observed even though many of the nodal papers involved, antedate the 1961 Science Citation Index by many years. "Important" work continues to be well cited long after its publication.

(2) Secondary authors of nodal papers were themselves highly cited in the *1961 Science Citation Index* (as primary authors of other papers) but were cited less than half as frequently (41.6 citations/author) as senior investigators.

(3) The above confirms a general impression that senior investigators are first authors for their major works. In our study, even the total number of citations (1,706) to all the nodal co-investigators is only 32% of all citations (5,329) to Asimov-distinguished senior nodal investigators.

(4) The chronological position in the *1961 Science Citation Index* of an author's nodal paper relative to his other cited works indicates that senior nodal authors are well "established" and coauthors to a lesser degree by the time the nodal papers are published.

(5) The citations in the *1961 Science Citation Index* to the total authorship of the nodal papers include only about one-third the number of self-citations attributed to the average author in the base file.

(6) The bulk (96%) of the total citations in the 1961 Science Citation Index to nodal authors was by non-nodal authors This fact demonstrates that the works of these nodal authors are in the mainstream of science and do not constitute a completely esoteric subgroup of papers. However, we note here the opportunity of developing a quantitative measure of the degree to which the works of a group of authors constitute a clique or "in group." For instance, there are 89 unique authors involved in the nodal papers in this study. There are a total of 57,800 unique primary source authors in the 1961 Science Citation Index. The nodal authors therefore constitute 0.154% of the source authorship in the index. Nodal authors appear as primary citing source authors 304 times as having cited nodal reference authors. The total number of citations to nodal reference authors was 7,035; thus, there were 4.32% of intragroup citations to all the works of nodal authors. The fraction of "in group" citations divided by the fraction of total authors (4.32/0.154 = 28.0) may be used as a simple approximation of the degree of citation cliquishness. This value should be about one if a given group of authors were engaged in random mutual citation.

(7) The average number of authors per nodal paper (2.15) is not significantly different from the average authorship reported for all biomedical papers. The proportion of nodal

papers with only one author (16/65) also was indistinguishable from reported averages.

(8) Evidence is presented demonstrating a citation leapfrogging effect across a span of many years. This effect may merely indicate an awareness by nodal authors of related work but may also constitute objective evidence for the idea that scientific achievements depend on previous advances. The frequency with which nodal authors are involved as references in the citation leapfrogging is plotted against the nodal paper numbers in a histogram. There is a sharp increase of involvement in citation leapfrogging that begins

with Watson and Crick whose nodal paper (27), published in 1953, advances an important theory of nucleic acid structure and may mark the coalescence of a new field of study, the molecular biology of the genetic code.

(9) Nodes 1, 2, 3, 6, 7, 10, 11, 14, 19, 23 and 26 highlight what we would subjectively consider to be the key breakthroughs which present new fundamental information in the evolving field or carry over vital information from other disciplines. The network that can be derived from an account by an historian highlights those events which that historian considers as fundamental. The lack of backreaching historical reference made evident by the drawing of an historical network facilitates a re-evaluation of the historian's assumption of fundamentality. In addition to a subjective re-evaluation, one may try to confirm or contradict the assumptions of fundamentality by looking for citation linkages from these "Fundamental Papers" back to other nodal works.

(10) It has been demonstrated that the nodal *work* of nearly fifty per cent of the recent (1941-1961) investigators was the most heavily cited work in the *1961 Science Citation Index* for the investigator who was first author. If articles which were the second most heavily cited work were included, the figure would increase to seventy per cent. Therefore, there may be value in using citation indexing as a tool for identifying those works by an author which are of historical significance. In nearly every exception to the above correlation, the most cited work post-dated the nodal work. This gives the impression that a later work (presumably on the same subject) provided a broader, more useful description of the nodal work and therefore is more often cited,

Citation Indexing of Nodal Bibliographies (NCI) Revealed the Following Facts :

(I1) In twenty-six instances, non-nodal authors were cited by three or more different nodes. Half of the 26 investigators were cited in the *1961 Science Citation Index* more heavily than the mean for senior nodal authors and 25 of the 26 were cited more heavily than the mean for junior nodal authors. The well-cited works of 4 of the 26 non-nodal authors were examined disclosing at least one new paper worthy of inclusion in the historical network. The historian might therefore profit by similar considerations for nodal citation indexes which can be created for histories of other scientific topics.

(12) Fifty-five per cent of the nodal research was performed in the United States.

(13) There were no appreciable number of extramural Public Health Service grants earlier than about 1946. Only the work involved in the later nodes (nodes 21-40) therefore could have been supported by P.H.S. funds. These 20 nodes involved 40 papers. Of these, twenty nodal papers (involving nine distinct nodes) explicitly acknowledge P.H.S. support. (See Appendix V.) In addition, Dintzis (Node 37) had a P.H.S. grant at the time of the work of his nodal paper though it was not acknowledged.

Further, one of the authors, Eisenstadt, involved in Node 38, had a P.H.S. fellowship at the time. Node 38 involves, however, three different papers. Furthermore, the research covered by three papers by Matthaei and Nirenberg in Node 40 were done at NIH in Bethesda. Therefore, 12 of the 20 nudes which postdate 1946 were supported to some extent by U.S.P.H.S. This support involved 27 of the 40 papers comprising these nodes. Thus the U.S. Public Health Service supported about two-thirds of the appropriate recent nodal work.

(14) This report also demonstrates a 65% coincidence between historical dependencies and the most straightforward citational dependencies. There are many instances where additional non-coincidental citation relationships exist between nodes.

(15) It is felt that citation analysis has been demonstrated to be a valid and valuable means of creating accurate historical descriptions of scientific fields, especially beyond the first quarter of the twentieth century when bibliographic citation had become well established as part of scientific publication.

#### APPENDIX I

#### SYNOPSIS OF THE BOOK, "THE GENETIC CODE" BY ISAAC ASIMOV

#### INTRODUCTION\*

In the history of science certain key discoveries, often based on a single profound observation, have opened the way to even greater strides in scientific knowledge. One such discovery was made by Avery et al (20) in 1944. They observed that deoxyribonucleic acid (DNA) carried genetic information which was capable of transforming one strain of bacteria to another different strain, that is, the strain from which the DNA was extracted. This brief story of the genetic code will attempt to explain the significance of Avery's discovery for the field of biochemistry, genetics, and molecular biology. CHAPTER I

For centuries man was cognizant of only the very obvious features of inheritance. Gregor Mendel (2) in the 1860's first demonstrated the predictability of dominant and recessive traits in plants, and thereby established the first laws of inheritance. Late in the 19th century histologists also studied the phenomenon of mitosis by which a cell, through division, is able to produce a replica of itself. In 1880 Walther Flemming (4) described the replication of paired chromosomes within the cell nucleus which preceded each mitotic division. Each new cell after division contained the same number and type of chromosomes possessed by the original cell. This constancy of chromosome replication throughout lifelong somatic cell division provided some indication that the chromosomes could carry information which determined the properties of each new generation of cells. The role of unpaired chromosomes in germ cell maturation and fertilization provided further evidence that the chromosome was the site of genetic information. The chromosome contains strings of genes. Each gene govems or specifies a particular characteristic of the future organism. The concept that spontaneous alteration of the chromosome can endow the organism with mutant characteristics was first expressed by Hugo de Vries (7) in 1900. OHAPTER II

The chromosome is largely protein in nature and is conjugated to nucleic acid (nucleoprotein). Nucleic acid was first isolated by Friedtick Miescher (3) in 1869. However, until recently, biochemists believed that genetic information was carried by the protein component of the chromosome. In 1935 Wendell Stanley (14) isolated crystals of tobaccomosaic virus. The virus, a parasitic invader of the cell, is able to replicate itself within

\*(Numbers in parenthesis are node designations). Authors in parenthesis are those not mentioned by Asimov, but who were identifiable by other descriptors. They are considered as senior nodal authors.

the cell as does the chromosome. In 1936 (Bawden & Pirie) (16) discovered that the virus, was also nucleoprotein. Therefore, by 1940 it was known that two different nucleoprotein entities were capable of replication.

#### CHAPTER III

A review of basic organic chemistry.

#### CHAPTER IV

Proteins, long considered the "stuff of life", are macromolecules consisting of chains of component amino acids. Braconnot (1) in 1820 was the first to isolate specific amino acids from protein. Any or all of twenty-two amino acids, occurring in any number or sequence, form the building blocks of a virtually unlimited variety of proteins. Emil Fischer (8), between 1900-1910, demonstrated the peptide chemical linkage of chains of amino acids forming a protein.

#### CHAPTER V

The structural description of protein must account for: (1) Its amino acid components and their sequence; (2) Its bending due to the formation of weak hydrogen bonds between segments of the polypeptide chain, and (3) The precise folding of the chain in space.

Attempts at determining the amino acid sequence of various proteins met with failure for many years. However, Martin and Synge (19) in 1944 developed the method of paper chromatographic separation of amino acids which provided a convenient means for isolation and analysis of protein components. Using this technique and a method of partial fractionation, Frederick Sanger (24) by 1953, was able to determine the amino acid sequence of insulin. Vincent Du Vigneaud (28) used Sanger's technique to determine the amino acid order of two other protein molecules, oxytocin and vasopressin; however, he proceeded one step further by synthesizing these proteins from the necessary amino acids.

Each type of protein formed by the organism is reproduced faithfully from specific types and numbers of amino acids, and in an inflexible order. This presumes a set of coded instructions which allows only select protein construction -- not randomization. CHAPTER VI

The chromosome seemed endowed with the blueprint for protein manufacture. Possible alteration of the chromosome by artificial means seemed the method of choice for studying this characteristic. Herman Muller(10), as long ago as 1926, was able to produce altered genes and mutants with x-rays. Beginning in 1941 Beadle and Tatum (18) subjected bread mold to X-rays and succeeded in producing mutant molds which required precise amino acid supplementation to the normal growth culture media of sugar and salts. They demonstrated that the X-rays altered a specific mold gene which controlled the manufacture of a specific enzyme (protein) used by normal mold to manufacture the amino acid from unsupplemented media. This assumption led to the one-gene-one-enzyme theory. Belief persisted that the gene might contain a reference protein (protein code) which was in fact the same as the protein (or enzyme) whose production was controlled by the gene. However, this reference

protein was never demonstrated nor was the existence of the complete series of 22 amino acids, common in tile adult, ever demonstrated in totipotential germ cells.

In 1928 it was shown (by Frederick Griffith) (1 I) that a strain of dead *capsulated* pneumocoeci, added to a culture of *living* non-capsulated pneumocoeci, could bring about the production of *living capsulated* bacteria. In 1931 (Alloway)(13) it was possible to achieve this transformation with an extract of the dead capsulated bacteria; therefore conelusive proof was presented that genetic material from a dead strain was influencing the characteristics of a live strain. Refinements of this genetic extract were sought until 1944 when Avery, Mac Leod and McCarty (20) identified the extract as protein-free DNA. This work conclusively proved that the genetic code could be carried by nucleic acid alone -- a fact whose impact would influence many disciplines of the life sciences.

Investigations turned to the phenomenon of replication of the virus. In 1952 Hershey and Chase (25) used tagged tracer methods to show that only the nucleic acid portion of bacteriophage virus entered the cell -- not the protein shell. However, while within the cell, the virus replicated itself many times over as a complete entity (nucleic acid and protein shell). This proved that: (1) nucleic acid, even from a virus, was able to replicate itself, and (2) that the viral nucleic acid was able to utilize the native amino acids within the cell to create a protein (the viral shell) foreign to the cell. In 1955 Fraenkel-Conrat (31) was able to separate the nucleicacid and protein shell of tobacco-mosaic virus. The nucleic acid by itself showed little infectivity to tobacco leaf; however, when recombined with its protein shell the virus again became infective. The protein therefore served as a protective capsule to the essential nucleic acid. These discoveries left no doubt that nucleic acid did indeed carry the genetic code.

#### CHAPTERVII

Fortunately, much of the chemical groundwork was in progress for over half a century prior to the revelation that DNA alone carried the genetic code. The purine and pyrimidine content of nucleic acid was studied by Kossel (5) and others during the 1880's. About 1910 Phoebus l,evene (9) identified the five carbon sugar ribose as the carbohydrate component of nucleic acid (Ribonucleic acid, RNA). Ribose had previously been isolated and synthesized by Emil Fischer (6) as a freely occurring sugar. Later l,evene (12) discovered that certain nucleic acids contained deoxyribose (DNA). Nucleic acid therefore contained either ribose or deoxyribose exclusive of all other sugars. The combination of (1) purine (adenine or guanine) or pyrimidine (thymine (only in DNA), uracil (only in RNA) or cytosine); (2) rihose or deoxyribosc, and (3) an attached phosphate group, was called a nucleotide. Levene (12) theorized that four of these nucleotides, each characterized by a different purine or pyrimidine group, formed nucleic acid (tetranucleotide theory). Levene (15) later proposed formulas which assigned definite linkages between the nucleotides. These were confirmed through chemical synthesis by Alexander Todd (29) in the early 1950's.

#### CHAPTER VIII

Levene's concept that only four nucleotides formed the nucleic acid molecule was based on crude methods of chemical separation of these entities. Milder extraction methods were used in the 1940-50 period and it became evident that a nucleic acid molecule (or the gene) might be formed of a chain of up to two thousand nucleotides. The demonstration by Avery et al (20) that DNA could carry genetic information made biochemists realize that the tetranucleotide hypothesis was invalid. The Martin and Synge discovery (19) of paper chromatography gave nucleic acid chemists the tool they required to properly analyze the makeup of nucleic acid. Erwin Chargaff (21), by 1947, demonstrated that purines and pyrimidines were present in unequal quantities within nucleic acids; also the ratio of one nucleotide to another differed from one nucleic acid to another. By the early 1050's Chargaff (22) was able to demonstrate that the different nucleotides in the chain were in random order. Therefore they could exist in great varieties'of combinations -- at least a sufficient enough number to determine a code for the amino acid order and content of hundreds of thousands of different proteins.

Watson and Crick (27) in 1053 employed X-ray diffraction methods for studies of nucleic acid. These methods were developed by Wilkins (26). They were able to construct a model of the spatial molecular configuration of DNA. This consisted of an interlocking helical arrangement of two polynucleotide chains about the same axis. The helical arrangement of polynucleotide chains had been considered a distinct possibility since Pauling and Corey (23) in 1951 presented the concept that polypeptide chains (of protein) could arrange themselves in a helical configuration through hydrogen bonding. The Watson-Crick model of DNA helped verify previous chemical data and, furthermore, provided a basis for understanding the replication of DNA on a molecular level. CHAPTER IX

The hydrogen bonding of the polynucleotide strands of the double helix exists at the position of a purine-to-pyrimidine approximation of the two strands. In DNA the purine adenine (A) will always attach to the pyrimidine thymine (T) (however in RNA uracil replaces thymine); further, the purine guanine (G) will always join the pyrimidine cytosine (C). Therefore, any approximate portions of the two strands are opposite and complementary (A-G-T-C vs. T-C-A-G). When the strands separate, each will act as a model for the recreation of the original complementary strand from individual nucleotides. Thus replication can be explained on a molecular basis.

Scientists sought to control methods of biochemical synthesis of nucleic acid. Severo Ochoa (32) in 1955 isolated a bacterial enzyme which produced polynucleotide strands of an RNA variety from adenosine diphosphate. Arthur Komberg (33) in 1956 produced synthetic polynucleotides of a DNA type from an enzyme, various deoxynucleotides and a

DNA "priming" strand. (The work of Ochoa and Kornberg closely approximated each other in time and scope. Both shared the 1959 Nobel prize. It is the only instance in the net-work diagram where each man is cited by the other.)

#### CHAPTER X

Experiments dating back to the early 1940's have shown that invariably the RNA concentration is highest in cells when the rate of protein synthesis is highest (1938 study by Caspersson and Schultz) (17). However, DNA is found only in the nucleus. Most of the RNA is contained in the cytoplasm (the site of protein synthesis), except for a small amount in the nucleus, which is that RNA most recently formed by the DNA of the nucleus. The code from a particular gene (DNA) forms a specific RNA which reaches the cytoplasm to control production of a specific protein. The DNA in this sense is the ultimate prototype of the protein.

The electron microscope and ultra cell centrifugation rnethods permitted investigation of the cytoplasmic microsomes which were rich in RNA and proved to be the site of amino acid incorporation into protein.

In 1953 George Palade (30) distinguished yet smaller particles associated with the microsomal fraction. He later isolated these particles or ribosomes and found they contained all the RNA in the microsomal fraction of the cell together with an equal amount of protein. Ribosomal RNA is therefore the exact site of protein synthesis but it does not carry the coded genetic instructions of DNA; rather it is the structural backbone, the "key blank", as it were, that could be impressed into service if it could be modified by a second RNA which *does* receive the imprint of the genetic code from DNA. The existence of this second RNA (Messenger RNA) was concluded in 1960 from investigation of bacterial cells (Jacob and Monod) (35). Messenger RNA was isolated from mammalian cells by Mirsky and Allfrey (39) in 1962.

#### CHAPTER XI

The genetic code consists of trinucleotide combinations or "triplets" running the length of the polynucleotide chain with each triplet representing a particular amino acid. Since there are 64 triplet possibilities and only 22 amino acids, some amino acids may be represented by more than one triplet. Therefore the code is said to be "degenerate". The triplet code does not overlap.

Mahlon Hoagland (34) in the late 1950's discovered that amino acids were combined with adenylic acid in an energy rich combination ("activated amino acid") before being incorporated into the polypeptide chain. Hoagland demonstrated a third type of RNA (freely soluble as short strands in the cytoplasm) which he termed Transfer RNA. Each strand of Transfer RNA consisted of a particular triplet with a code affinity to a particular type of activated amino acid. These combine and attach to a specific position on Messenger RNA where a complementary triplet exists. Dintzis (37) in 1961 demonstrated that this concept of protein construction was accurate. He demonstrated that all the amino acids in a

molecule of hemoglobin could be set in place and bound together in a mater of 90 seconds. The whole scheme was duplicated in a laboratory with the use of cell fragments. In 1961, Hurwitz (36) used a system of DNA, nucleotides, and enzymes and succeeded in manufacturing Messenger RNA in a test tube. Novelli (38) in 1961 carried the process one step further by using DNA nucleotides and also ribosomes and amino acids. He succeeded in manufacturing Messenger RNA which in turn coated the ribosomes. This combination acted as a model for the formation of a particular protein, the enzyme, beta-galactosidase.

The ultimate verification of the triplet code theory came in 1961 when Nirenberg and Matthaei (40), using Ochoa's synthetic method, formed a polynucleotide containing just one polynucleotide, polyuridylic acid. This synthetic Messenger RNA thereby consisted of a chain of triplets with the code U-U-U. In a system containing a variety of amino acids a protein was formed which utilized only one amino acid -- phenylalanine. Therefore, the triplet U-U-U- meant phenylalanine. This discovery is the first step in the ultimate understanding of the genetic code. Its consequences will be left to future history.

#### APPENDIX II

#### DETAILED DESCRIPTION OF NODAL CITATION CONNECTIONS AND WEIGHTINGS IN THE NETWORK CHARTS

- METHOD A. Bibliographies of nodal articles were searched for citations to earlier nodal authors. The following methods of search were used to demonstrate relationships.
  - 1. Each bibliography was searched for *direct citation* of another nodal paper. Example: Smith 1960 to Jones 1940. (*Strong Direct*)
  - 2. Each bibliography was searched for citations to non-nodal papers by nodal authors which were published subsequent to the cited author's nodal paper. Example: Smith 1960 *through Jones* 1950 to Jones 1940 (*Strong Indirect*).
  - 3. The texts, footnotes, and bibliographies of nodal papers were searched for descriptions of earlier nodes in which a nodal author was acknowledged although no exact reference citation was given. (Weak Indirect). (When a more direct connection was established between two particular nodes, any less direct connection between the two nodes was ignored.)
- METHOD B: In a few instances the above methods did not provide connections leading *from* a node to any earlier node. In these instances the following methods were used.
  - 4. The bibliographies of nodal papers were searched for self-citations involving any nodal co-author including those not mentioned by Asimov. The bibliographies of these self-cited references were examined for citation to a prior node. Example: Smith 1960 through Smith •950 to Jones 1940. (Strong Indirect Self-Citation).
  - If this failed the following method was used. Each bibliography of *every* reference cited in the node article was searched for citations to earlier nodes. Example: Smith 1960 *through Brown 1950* to Jones 1940. (Weak Indirect.)

The term *strong* as applied to citation connections is used here to indicate a citation pathway established directly, or indirectly through use of intermediate papers by the same nodal authors.

The term *weak* as applied to citation connections is used here to indicate a citation pathway established through use of intermediate papers by non-nodal authors. The term weak also implies the use of incomplete citation data such as personal communication, incomplete text reference, etc. as a connecting link.

It should be carefully noted that the possible importance, in the total historical picture, of these non-nodal intermediates is not implied by the word "strong", nor is it denied by the use of the word "weak".

The procedure used in METHOD B above (using intermediate non-nodal authors,

## **Nodal Weighting Values**

An arbitrary weighting factor is assigned each node as an expression of the strength of total citational connections of the node. This binary term is calculated as the sum of the weights of each citational connection entering or leaving the node. A strong direct citation (solid blue lines, 3rd overlay from the bottom, and solid yellow lines, 5th overlay from the bottom) is given a value of 4, a strong indirect citation (broken lines 3rd and 5th overlays) is given a value of 2, and a weak indirect citation (solid or broken blue lines, 4th overlay from the bottom, and solid or broken yellow lines 6th overlay from the bottom) is given a value of 1. The nodal articles are ranked in the following list wherein the paper by Devries (node 7) has the lowest value (00000), and the paper by Avery (node 20) has the greatest nodal weighting (110112 = 2710). The same nodal value is assigned each article in cases when the node is composed of more than one article.

NODAL			TYPE			
VALUE	FIRST AUTHOR	PUBLICATION	OF PAPER	YEAR	VOL.	PAGE
			/			
DEVRIE	ŚН	CR AC SCI	-{	00	130	845
BRACON	NOT H R E	AN CHIM P Z AN CHEM	 M	20 07	13 20	113 913
BEADLE DUV IGN DUV IGN MENDEL MULLER PAUL IN PAUL IN PAUL IN	GW EA•V EA•V G HJ G L G L G L	PNAS JACS JACS VERHNAT BRJEXB JACS PNAS PNAS	- -L -L -R -L -	41 53 53 65 26 50 51 51	27 75 10 72 37 37	499 4879 4880 85 5349 205 235
DINTZI FISCHE FLEMMI HERSHE WILKIN WILKIN	S HM R E NG W Y AD S MHF S MHF	P N A S BER DTSCH ARC MIK A J G PHYSL B B ACTA NATURE	- - -L	61 91 79 52 53	47 24 36 10 171	247 4214 302 39 192 738
MICHEL	SO. AM	J CHEM S	-	55		2632
KÖŠSEL LEVENE PALADE PALADE	A PA GE GE	Z PHYSL C BER CHEM BER CHEM J B B CYT J EX MED	-	86 09 56 54	10 42 42 100	248 2102 3247 171 641
HÖAGLA HÖAGLA SANGER SANGER SANGER	ND MB ND MB F F	B B ACTA J B C BIOCHEM J BIOCHEM J BIOCHEM J BIOCHEM J	-L - -	57 58 51 51 53 53	24 231 49 53 53	215 241 463 481 353 366
	Y 11	L EV MED		22	55	01

CONSDEN R GORDON AH GRIFFITH F LEVENE PA STANLEY WM	BIOCHEM J- BIOCHEM J-M J HYGIENE- J B C - SCIENCE -L	443 28 35 35	38 22 37 R 1 27 11 109 62 81 64	43334
FRAENKEL • H FRAENKEL • H FRAENKEL • H	BBACTA - JACS -L PNAS -	57 56 55	25 8 78 88 41 69	87 82 90
EISENSTA.JM JACOB F KAMEYAMA T KORNBERG A KORNBERG A NOVELLI GD SIBATANI A	P N A S - J MOL BIO-R P N A S - B B ACTA -L FED PROC -M JHU MCP I- SCIENCE -L P N A S -	62 61 62 56 56 57 61 62	48 65 3 31 48 65 21 19 153 57 133 136 48 47	289771991
BAWDEN FC BAWDEN FC CASPERSS.T CASPERSS.T	NATURE −A PRSBIDL÷ NATURE →L NATURE −L	36 37 38 39	138 105 123 27 142 29 143 60	14
LEVENE PA LEVENE PA MATTHAEI JH MIESCHER F NIRENBER.MW NIRENBER.MW WATSON JD WATSON JD	J B C - J B C - P N A S - H-S M C U- P N A S - P N A S - NATURE - NATURE -	29 29 61 71 62 53 53	83 79 83 80 47 158 460 44 47 158 48 10 171 73 171 96	330118474
HURWITZ J	B B RES C-	60	3 1	5
CHARGAFF E	C SPR H S-M	47	12 2	8
CHARGAFF E	EXPERIENT-T	50	6 20	1
GRUNBERG.M GRUNBERG.M DCHDA S	JACS-L SCIENCE - FED PROC -	55 55 56	77 316 122 90 15 83	572
AVERY OT	J EX MED -	44	79 13	7

or self-citation pathways) was not employed when a citation line to any earlier node could be established by means used in METHOD A above. It is obvious, therefore, that other citation lines could be established by investigating *all* self-citations and all other references as possible citation intermediates. The use of the more exhaustive METHOD B could not economically be applied to all the papers in the study.

Only the methods used above are displayed on the Network Charts. NODE VALUES

Arbitrary weighting values were assigned the above connections.

CONNECTION	WEIGHT
Direct	4
Strong Indirect	2
Weak Indirect	1

Using these weights, each node can be assigned a value (expressed as a binary number) depending on the number and type of connections which enter and leave it. (In instances in which a node is composed of two or more papers, each source paper is assigned the value for the composite node.)

An example of calculating a nodal weight is given below:

Node 20(Avery et al) is cited directly by three nodes and indirectly by one node. Node 20 directly cites two nodes and cites three other nodes indirectly. Therefore, nine connecting lines are associated with the node.

DIRECT LINES, 5

(weight x 4) 20

**INDIRECT LINES**, 4

Breakdown - STRONG INDIRECT, 3 (weight x 2)	6
WEAK INDIRECT, 1 (weight x 1	1
TOTALNINI	07

TOTAL Node Value ...... 27

#### NODAL CITATION RELATIONSHIPS

In the following listing, relationships demonstrated by literature searching methods for each node are exactly described. The intermediate references used as pathways between nodes are listed. Referral to the Network Charts will orient the reader.

Node 40 Nirenberg and Matthaei 1961-62

- A. Recent end point of study therefore not cited.
- B. Direct citation to Hurwitz (36).
- C. Strong indirect citations.
  - 1. Kirsch, Siekevitz, & Palade: J. Biol. Chem. 235:1419 1960 to Palade (30).

(Number in parenthesis is the nodal number.)

- 2. Hoagland: *Proc. Nat. Acad. Sci. U.S.* 46:1554 1960 to Hoagland: Proc. 4th Int. Congress Biochem. VIII.Vienna 1958 to Hoagland (34).
- 3. Hershey: J. Gem Physiol. 38:145 1954 to Hershey: J. Gen. Physiol. 37:1 1953 and Hershey, Dixon and Chase: J. Gen. Physiol. 36:777 1952 to Hershey and Chase (25).
- D. WeakIndirect
  - 1. Personal communication to Ochoa (32).
  - 2. Personal Communication to Fraenkel-Conrat (31)
- Node 39 Allfrey and Mirsky 1962
  - A. Recent end point of study therefore not cited.
  - B. Direct Citation to Hurwitz (36), to Jacob, & Monod (35).
  - C. Strong Indirect Citations
    - 1. Hoagland in "Nucleic Acids" 1960, vol. 3, pg. 360 to Hoagland (34).
- Node 38 Novelli 1961-62
  - A. Recent end point of study therefore not cited.
  - B. Direct citation to Hurwitz (36), to Jacob & Monod (35).
  - C. Strong Indirect citation:
    - 1. Ochoa: Proc. Nat. Acad. Sci. U.S. 47:670 1961 to Grunberg-Manago, Ortiz & Ochoa: Biochim. et Biophys. 20:269 1956 to Ochoa (32).
- Node 37 Dintzis 1961
  - A. Recent end point of study therefore not cited.
  - B. No direct citations.
  - C. No strong indirect citations.
  - D. Weakindirect citations:
    - 1. Steinberg et al: Science 124:389 1956 to Sanger (24), to Ochoa (32).
    - 2. Loftfield & Eigner: J. Biol. Chem. 231:925 1958 to Hoagland (34).
    - 3. Loftfield, Proc. 4th Int. Congress Biochem. VIII. 222 1960 to Hoagland (34).
    - 4. Borsook: Proc. 3rd Int. Congress Biochem., p. 92 1956 to Caspersson (17).
    - 5. Osawa & Satake: J. Biochem., (Tokyo) 42:641 1956 to Sanger (24).

Node 36 Hurwitz 1960

- A. Cited by (38)(39)(40).
- B. No direct citations.
- C. No strong indirect citations.
- D. Weakindirect citation.
  - 1. Weiss & Gladstone, J. Am. Chem. Soc. 81:4118 1959 to Ochoa (32).
- Node 35 Jacob and Monod 1960-61
  - A. Cited by(38) (39).
  - B. No direct citations.
  - C. Strong indirect citations.
    - 1. Kornberg et al: Proc. Nat. Acad. Sci. U.S. 45:772, 1959 to Kornberg (33).

Node 34 Hoagland 1957-58

- A. Cited indirectly by (37) (39) (40).
- B. No direct citations.
- C. Strong indirect citations.
  - 1. Caspersson: Cell Growth and Cell Function, N.Y. 1950 to Caspersson (17).

Node 33 Kornberg 1956-57

- A. Cited by (32); cited indirectly by (35).
- B. Direct citation to Ochoa (32).
- C. No strong indirect citations.

Node 32 Ochoa 1955-56

- A. Cited by (33); cited indirectly by (36) (37) (38) (40).
- B. Direct citation to Kornberg (33) to Watson & Crick (27),to Fraenkel-Conrat (31).
- C. Strong Direct citation.
  - 1. Vischer & Chargaff: J. Biol. Chem. 176:715, 1948 to Chargaff (21).
- D. Weak indirect citation.
  - 1. Descriptive text reference to Todd (29).

Node 31 Fraenkel-Conrat 1955-57

- A. Cited by (32); cited indirectly by (40).
- B. No direct citations.
- C. Strong indirect citations.
  - Cohen & Stanley: J. *Biol. Chem.* 142:863 1942 to Stanley & Loring: Cold Spr. Har. Sym. 6:341 1938 and Loring & Stanley: J. Biol. Chem. 117:733 1939 to Stanley (14).
  - 2. Holden & Pirie: Biochem J. 60:46 1955 to Bawden & Pirie (16).
- Node 30 Palade 1954-56
  - A. Cited indirectly by (40).
  - B. Direct citation to Avery et al (20)
  - C. No strong indirect citations.
- Node 29 Todd 1955
  - A. Cited indirectly by (32).
  - B. No direct citations.
  - C. Strong indirect citations.
    - 1. Michelson & Todd: J. Chem. Soc. p. 34 1954 to Levene (15).
    - 2. Dekker, Michelson & Todd: J. Chem. Soc. p. 947 1953 to Levene (12).
- Node 28 DuVigneaud 1953
  - A. Not cited.
  - B. No direct citations.
  - C. Strong indirect citation.
    - 1. Popenoe & DuVigneaud J. Biol. Chem. 205:133, 1953 to Sanger (24).

Node 27 Watson & Crick 1953

- A. Cited by (32).
- B. Direct citation to Wilkins (26).
- C. Strong indirect citations.
  - 1. Pauling & Corey: Proc. Nat. Acad. Sci. U.S. 39:84 1953 to Pauling (23).
  - 2. Zamenhof, Bawerman & Chargaff: *Biochim. et Biophys.* 9:402, 1953 to Chargaff (22).
- Node 26 Wilkins 1953
  - A. Cited by (27).
  - 13. No direct or indirect citations.
- Node 25 Hershey and Chase 1952
  - A. Cited indirectly by (40).
  - B. No direct citations.
  - C. No strong indirect citations.
  - D. Weak indirect citations.
    - Anderson: Botany Rev. 15:464 1949 cites both Stanley & Anderson J. Biol. Chem. 139:325 1941 to Bawden & Pirie (16) and Muller H.J. Proc. Roy. Soc. Lond. (B) 134:1 1947 to Avery et al (20).
- Node 24 Sanger 1951-53
  - A. Cited indirectly by (28) (37).
  - B. Direct citation to Martin & Synge (19).
  - C. No strong indirect citations.
- Node 23 Pauling and Corey 1950-51
  - A. Cited indirectly by (27).
  - 13. No direct or indirect citations.
- Node 22 Chargaff 1950
  - A. Cited indirectly by (27).
  - B. Direct citation to Martin and Synge (19), Avery et al(20)Chargaff (21).
  - C. Strong indirect citation.
    - 1. Tipson: Adv. Carbohydrate Chem. 1:193, 1945 to Levene & Tipson (15).
- Node 21 Chargaff 1947
  - A. Cited by (22); indirectly cited by (32).
  - B. Direct citation to Avery et al (20), Miescher (3).
  - C. No strong indirect citations.
- Node 20 Avery, MacLeod and McCarty 1944
  - A. Cited by (30) (22) (21). Cited indirectly by (25).
  - B. Direct Citation to Alloway (13), Griffith (11).
  - C. Strong indirect citations.
    - 1. Levene & Dillon: J. Biol. Chem. 96:461 1933 to Levene (12).
    - 2. Schultz: Cold Spr. Har. Sym. 9:55, 1941 to Caspersson & Schultz (17).

3. Stanley: Handbuch der Virusforschung 1:491 1938 to Stanley (14).

(This node (20) is considered the major breakthrough by Asimov. In the citation diagram it has the highest number of connecting lines and the highest node value).

Node 19 Martin and Synge 1943-44

- A. Cited by (24) (22).
- B. No direct or indirect citations.
- Node 18 Beadle and Tatum 1941
  - A. Not cited.
  - B. No direct citations.
  - C. Strong indirect citation.
    - 1. Sturtevant & Beadle: An Introduction to Genetics 1931 to Mendel (2).
- Node 17 Caspersson and Schultz 1938-39
  - A. Cited indirectly by (37) (34) (20).
  - B. Direct citation to Bawden and Pirie (16).
  - C. Strong indirect citation.
    - 1. Muller: J. Genet. 22:229 1930 to Muller (i0).
- Node 16 Bawden and Pirie 1936-37
  - A. Cited by (17); cited indirectly by (30) (25) (20).
  - B. Direct citation to Stanley (14).
  - C. No strong indirect citations.
- Node 15 Levene and Tipson 1935
  - A. Cited indirectly by (29) (22).
  - B. Direct citation to Levene (12).
  - C. No strong indirect citations.
- Node 14 Stanley 1935
  - A. Cited directly by (16); cited indirectly by (31) (20).
  - B. No direct citation to node.
  - C. No indirect citations.
- Node 13 Alloway 1932
  - A. Cited by (20) .
  - B. Direct citation to Griffith (11).
  - C. No strong indirect citations.
- Node 12 Levene with Mori and London 1929
  - A. Cited by (15); cited indirectly by (29) (20).
  - B. No direct citations.
  - C. Strong indirect citations.
    - 1. The "work of Kossel" as described in Jones W: Nucleic Acid 2nd ed., New York, p. 136, 1920 *to* Kossel (5).
    - 2. Levene & Jacobs; J. Biol. Chem. 12:411 1912 to Levene (9).

Node 11 Griffith 1928

A. Cited by (20) (13).

B. No direct or indirect citation to node.

- Node 10 Muller 1926
  - A. Cited indirectly by (17)
  - B. No direct or indirect citations.
- Node 9 Levene and Jacobs 1909
  - A. Cited indirectly by (12).
  - B. Direct citation to Fischer & Piloty (6).
  - C. No strong indirect citations.
- Node 8 Fischer 1907
  - A. Not cited.
  - B. No direct citations.
  - C. No strong indirect citations.
  - D. Weak indirect citation.
    - 1. Descriptive text reference to Braconnot (1).
- Node 7 DeVries 1900
  - A. Not cited.
  - B. No direct or indirect citation (no references).
- Node 6 Fischer and Piloty 1891
  - A. Cited by (9).
  - B. No direct or indirect citation\*.
- Node 5 Kossel 1886
  - A. Indirectly cited by (12).
  - B. Direct citation to Miescher (3).
  - C. No direct or indirect citation\*.
- Node 4 Flemming 1879
  - A. Not cited.
  - B. Direct citation to Miescher (3).
  - C. No strong indirect citation\*.
- Node 3 Miescher 1871
  - A. Cited by (21) (5) (4).
  - B. No direct ot indirect citation. This paper represents an origional work, that is, the discovery of nucleic acid.
- Node 2 Mendel 1865
  - A. Indirectly cited by (18).
  - B. No direct or indirect citation

Bateson states that Focke provides the only instance before 1900 in which Mendel was cited. He states that Mendel's work was rediscovered by DeVries (Node 7), Correns and

\*Papers listed in the node bibliography were not investigated to determine if weak indirect connections existed, because of the difliculty of procuring foreign references over 70 years old.

Tschermarr in 1900. [Bateson W: Mendel's Principles of Heredity, Cambridge Univ. Press, 1909, p. 317-361; Focke: Pflanzewimschlinge, p. 109, 1881.]

Node 1 Braconnot 1820

- A. Indirectly cited by (8).
- B. No direct or indirect citations. (Original work, earliest node).

Non-Connective Citations to Nodal Authors

In certain nodal bibliographies, citations were made to early nodal *authors*, the cited work being more recent than *paper(s)* comprising the node. However, these cited references did *not*, in these instances, provide strong indirect connections between nodes, i.e. they do not lead to the earlier *nodal papers*. Although the network chart does not indicate these cases; they are worthy of historical note.

1. (40) Nirenberg and Matthaei cite.

Tissieres, Watson, Schessinger & Hollongsworth, J. Mol. Biol. 1:221, 1959 which cites Tissieres & Watson, Nature 182:778, 1959 which does not cite Watson (27).

- Hurwitz (36) cites Rose, Grunberg-Manago, Corey and Ochoa, J. Biol. Chem. 211:737, 1954 which does not cite Ochoa (32).
- (33) Kornberg cites
   Brawerman & Chargaff: 1. Amer. Chem. Soc. 75:2020, 4113, 1953 which cites Vischer
   & Chargaff, J. Biol. Chem. 176:175, 1948 which does not cite Chargaff (21).
- 4. (31) Fraenkel-Conrat cites Watson, *Biochim. et Biophys.* 13:10, 1954 which does not cite Watson (27).
- 5. (17) Caspersson and Schultz cite Stanley, *Amer. Nat.* 62:110, 1938 which does *not* cite Stanley (14).

## APPENDIX III

# Citation Index Prepared from the 65 Nodal Papers (NCI)



## **APPENDIX IV**

## **Work Locations Specified By Nodal Articles**

An article often indicates the location where, or organization under which the investigation was conducted. From the nodal papers, twenty-five locations are listed together with the number of articles for each location. Since certain nodes contain multiple articles, the actual number of nodes represented for each location is also listed. The Rockefeller Institute for Medical Research was the location where the work constituting eight nodes was conducted, and therefore is most important in the historical scheme.

Z

IT UT IG	DES LVED			TYPE				NOVELLI GD SCIENCE -L 61 133	1369
NST NO.	NVO	FIRST		OF	YEAR	VOL.	PAGE	PHYSIOL I BERLIN 86 10	248
I /	- Т Т Е			1		ļ		*RUCKEF I MED RES	91 137 2102
1	PAU	LING L LING L LING L	JACS PNAS PNAS	-'L - -	5'0 51 51	72 37 37	5349 205 235	LEVENE PA BER CHEM - 09 42 LEVENE PA J B C - 29 83 8 LEVENE PA J B C - 29 83	3247 793 803
*CARN	HER	SHEY AD	J G PHYS	L-	52	36	39	PALADE GE J B B CYT- 56 2	171
*CARU 1 *COLU		PERSS.T	NATURE NATURE	-L -L	38 39	142 143	294 602	SIBATANIA PNAS – 62 48 STANLEY WM SCIENCE –L 35 81	471 644
+CORN	CHA	RGAFF E	EXPERIEN	T-T	50	6	201	1 BAWDEN FC NATURE -A 36 138 BAWDEN FC P RS BIOL- 37 123	1051
1		IGNEA.V	J A C S J A C S	-L -L	53 53	75 75	4879 4880	*STANFORD U	499
*HARV 1	HOA	GLAND M	B B B ACTA	-L	57	24	215	1 MIESCHER F H-S M C U- 71 460	441
*.I PA	STEU	R PARIS	J MOL BI	0-R	61	3	318	I FRAENKEL H B B ACTA – 57 25 1 FRAENKEL H J A C S –L 56 78	87 882
+KING	FLE	MMING W	ARC MIK	A-	79	16	302	*U CAMBRIDGE	1051
1	WIL	KINS ME	F B B ACTA	-L	53 53	$10 \\ 171$	192 738	BAWDEN FC PRS BIOL- 37 123 MICHELSD-AM J CHEM S - 55 SANGED F BIOCHEM L 51 40	274
- MA33	DIN	TZIS H	PNAS	-	61	47	247	4 SANGER F BIOCHEM J- 51 49	481
	GRI	FFITH F	J HYGIEN	E-	28	27	113	SANGER F BIOCHEM J- 53 53	366
1	MAT	THAEI J	H P N A S	-	61 61	47	1580	WATSON JD NATURE - 53 171	964
*NYU	COLL	EGE MEI				70	2165	•U WURZBURG	4214
2	GRU	NBERG.N	SCIENCE B B RES	c-	55	122	907	WASH U SCH MED	197
+OAKR	IDGE	NAT LA	B			15	632	KORNBERG A JHU MCP I- 57 15	579
1	KAM	EYAMA 1	PNAS	-	62	48	659	1 CONSDEN R BIOCHEM J- 44 38	3 224

## **APPENDIX V**

## **Agencies Supporting The Research**

Most nodal articles, especially those of recent years, list the contributing agencies which provided funds for the investigations.

The papers in node 32 (represented by first authors Grunberg-Manago and Ochoa) received the most diverse support.

The U.S. Public Health Service provided the most extensive support since it contributed to work forming *nine* nodes.

ES TING ARCH

5 K K							
AGEN UPPO IE RE		TYPE OF				PAULING L P N A S - 51 37 2 PAULING L P N A S - 51 37 2	05 35
↓ S ∓	∠ Z Z AUTHOR	PUBLICATION PAPER	YEAR	VOL.	PAGE	A TI ART MET DIS 1 GRUNBERG.M SCIENCE - 55 122 9 1 OCHOA S FED PROC - 56 15 8	07 32
AMER	GRUNBERG M	JACS -L	55	. 17	3165	NAT I HEALTH	15
3	HOAGLAND MB	J B C - B B ACTA -L	58 56	231	241	I KORNBERG A B B ACTA -L 56 21 1 KORNBERG A JHU MCP I- 57 153 5	97 79
	OCHOA S	FED PROC -	56	155	832	I JACOB F J MOL BIO-R 61 3 3	18
1 ATOM	DUVIGNEA.V DUVIGNEA.V ENERGY COM	JACS-L JACS-L	53 53	75 75	4879 4880	GRUNBERG.M JACS -L 55 77 31 1 GRUNBERG.M SCIENCE - 55 122 9 OCHOAS FED PROC - 56 15 8	65 07 32
1	HOAGLAND MB	B B ACTA -L J B C -	57 58	24 231	215 241	PUB HEALTH SERV	01
	JACOB F	J MOL BIO-R	61	3	318	FRAENKEL-H P N A S - 55 41 6	90
	HURWITZJ	B B RES C-	60	3	15	GRUNBERG M SCIENCE - 55 122 9	07
	CHARGAFF E	EXPERIENT-T	50	6	201	HOAGLAND MB B B ACTA -L 57 24 2	15
JANE 2	HURWITZ J JACOB F	B B RES C- J MOL BIO-R	60 61	3	15 318	HURWITZ J B B RES C- 60 3 KORNBERG A B B ACTA -L 56 21 1 KORNBERG A HU MCP I- 57 153 5	15
1	SANGER F	BIOCHEM J- BIOCHEM J-	51 51	49 49	463 481	DCHOLSS FED PROC - 56 15 8 PAULING L P N A S - 51 37 2 PAULING L P N A S - 51 37 2	32
	HERSHEY AD	J G PHYSL-	52	36	39	SIBATANI A PNAS - 62 48 4	7í
1	FRAENKEL .H FRAENKEL .H	B B ACTA - P N A S -	57 55	25 -41	87 690	BEADLE GW PNAS - 41 27 4 GRUNBERG.M JACS -L 55 77 31 CRUNBERG.M SCIENCE - 55 122 9	99
2	FRAENKEL .H FRAENKEL .H FRAENKEL .H PAULING L	B B ACTA - J A C S -L P N A S - J A C S -L	57 56 55 50	25 78 41 72	87 882 690 5349	3 OCHOA S FED PROC - 56 15 8 PAULING L J A C S -L 50 72 53 PAULING L P N A S - 51 37 2 PAULING L P N A S - 51 37 2	32 49 05 35

## **APPENDIX VI**

## **Index of Nodal Papers**



### Type of Paper

-	ARTICLE

- -L LETTERS & PRELIMINARY NOTES
- --M PAPER PRESENTED AT A MEETING --R REVIEW ARTICLE
- --T LECTURES

### **Publication Abbreviations**

AM NAT	AM NATURALIST		
AN CHIMP	ANN CHIM PHYS	JBC	J BIOL CHEM
ARC MIK A	ARCHMIKROSKOP ANAT	J CHEM S	J CHEM SOC
В В АСТА	BIOCHEM BIOPHYS ACTA	J EX MED	J EXP MEDICINE
B B RES C	BIOCHEMBIOPHYS RES COMMUN	J G PHYSL	J GENERAL PHYSIOL
BER DTSCH	BER DEUTSCH CHEM	J HYGIENE	J HYGIENE
BIOCHEM J	BIOCHEM J	J MOL BIO	J MOLEC BIOL
BR J EX B	BRIT J EXP BIOL	JHU MCP I	JOHNSHOPKINS U , MCCOLLUM-PRATT I
C SPR H S	COLD SPRING HARBOR SYMP	NATURE	NATURE
CR AC SCI	COMPTREND ACAD SCI	PNAS	PROC NAT ACAD SCI
EXPERIENT	EXPERIENTIA	P RS BIOL	PROCROY SOC BIOL
FED PROC	FED PROC	SCIENCE	SCIENCE
HS M C U	HOPPE-SEYLERS MED CHEM UNTERS	VERH NAT	VERHANDL NATURFORSCH
JACS	J AM CHEM SOC	Z AN CHEM	ZEITSCH ANGEW CHEM
J B B CYT	J BIOPHYS BIOCHEM CYTOL	Z PHYSL C	ZEITSCH PHYSIOL CHEM

#### APPENDIX VII

#### LEGEND FOR THE NETWORK CHARTS

#### ASIMOV'S CONNECTIONS RED OVERLAYS 1 & 2

*First* overlay (red) - Asimov's specified historical connections -- solid lines. *Second* overlay (red) - Asimov's implied historical connections -- broken lines.

#### COINCIDENT CITATION CONNECTIONS BLUE OVERLAYS 3 & 4

*Third* overlay (blue) - Coincident strong citation connections -- strong citation connections which *coincide* with Asimov's historical connections, specified and/or implied.

Blue solid line -- strong direct citation of one node by another.

blue heavy broken line -- strong indirect citation connection. These connections were determined by finding an intermediate paper by an earlier nodal author which was cited by a later nodal author.

Blue fine broken lines -- strong indirect citation connection established by finding an intermediate paper by the later nodal author which in turn cites the earlier nodal author.

Fourth overlay (blue) - Coincident weak citation connections -- weak citation connections also coincide with Asimov's description.

Solid line -- implied citation connection where a nodal author refers to the work of an earlier nodal author by text description or through personal communication but not by explicit citation.

Blue broken lines -- weak indirect citation connection established by one intermediate paper by a non-nodal author.

#### NON-COINCIDENT CITATION CONNECTIONS YELLOW OVERLAYS 5 & 6

*Fifth* overlay (yellow) - Non-coincident strong citation connections. Citation connections which do *not* coincide with Asimov's historical connections.

Solid line -- strong direct citation of one node by another

Broken line -- indirect citation connection where connections were determined by finding an intermediate paper by an earlier nodal author which was cited by a later nodal author.

Fine broken line -- indirect citation connection established by finding an intermediate paper by the later nodal author which in turn cites the earlier nodal author.

*Sixth* overlay (yellow) - Non-coincident weak citation connections. Citation connections which do *not* coincide with Asimov's historical connections.

Solid line -- implied citation connection where a nodal author refers to the work of an earlier nodal author by text description or through personal communication but not by explicit citation.

Broken line -- indirect citation connection established by one intermediate paper by a non-nodal author

COLOR CODES FOR COMBINATIONS OF TRANSPARENCIES

When all transparent overlays are combined or superimposed a complete comparative picture is observed -- both coincidence and non-coincidence of the Asimov historical network and citation network.

The nodes which were *not* reinforced by citation connections stand out as pure red lines. The citation connections which coincide with Asimov's historical connections are purple, that is, a combination of red and blue. The same information is revealed by examining the blue overlays separately.

Citation connections which are not coincident with Asimov's historical connections stand out as pure yellow lines.

The composite of all six overlays reveals those connections established by Asimov alone, by citation data alone, or a combination of the two.

A composite of the top four overlays (third through sixth) represents citation data. However, the reader should keep in mind that the citation connections are those established almost exclusively on the basis of nodal data, not on the basis of locating citation data from all possible sources.

Nodes are indicated by blocks assigned in chronological order. Each block contains the nodal number, nodal author named by Asimov, and the years covered by the nodal work. (Secondary authors are only included in nodes 6, 9, 12, 15 in order to distinguish these nodes from others in which Levene and Fischer are also involved.) The topological display of the nodes is organized so that nodes for broad fields are alligned together. Each broad field has a comer code indicated below:



In some nodes combinations exist. For example, Node 20 is coded both for bacterial genetics and nucleic acid chemistry.

Starting near bottom left one can see the development of protein chemistry. At the center the field of genetics is traced and on the right nucleic acid chemistry. One can see that the various fields coalesce as molecular biology towards the center and top of the network.

















