THE FLORIDA STATE UNIVERSITY

SCHOOL OF INFORMATION STUDIES

# PIGEONHOLES AND PUNCHCARDS: IDENTIFYING THE DIVISION BETWEEN LIBRARY CLASSIFICATION RESEARCH AND INFORMATION RETRIEVAL RESEARCH, 1952-1970

By

SHAWNE D. MIKSA

A Dissertation submitted to the
School of Information Studies
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded
Summer Semester, 2002

The members of the Committee approve the dissertation of Shawne D. Miksa defended July 3, 2002.

Kathleen Burnett
Committee Chair

Marcy P. Driscoll
Outside Committee Member

Jane Robbins
Committee Member

Gary Burnett
Committee Member

Approved:

Jane Robbins, Dean, School of Information Studies

## DEDICATION

To Heather

To Dad

"...these are only hints and guesses,
Hints followed by guesses; and the rest
Is prayer, observance, discipline, thought and action."

T.S. Eliot
(*Four Quartets*)

## ACKNOWLEDGMENTS

I owe a great deal of thanks to many people. First, to my committee. My advisor, Dr. Kathleen Burnett, for being a good friend, a mentor, for taking on all my neurosis, and for letting me vent in oddly hysterical ways. Dean Jane Robbins for knowingly, or unknowingly, admitting an opinionated know-it-all into the school and for allowing me to pour her drinks. Dr. Gary Burnett for taking over Elfreda's role, (thank you! thank you!), for a being a true Geek and for not giving me any slack. (I say that with the utmost respect. Really.) Dr. Marcy Driscoll for braving the unknown territory of library science. Dr. Elfreda Chatman for kindly, but brutally, saying "You can do better" when handing me back my first theory paper. Rest in peace you great big wonderful woman.

I owe a great debt to Dr. Katherine McCain and Dr. Howard White at Drexel University for allowing me to sit at their feet and learn the citation analysis and for just being encouraging and open and honest about this study. Also, thanks to Professor Ia McIlwaine and Professor Vanda Broughton, at the School of Library Archive and Information Studies, University College London, for letting me hang out and ask many questions. Also, Mr. Eric Coates, Mr. Jack Mills, and Mr. D.J. Foskett—it was wonderful to watch you three work.

I have many friends to thank. Dr. Laurie Bonnici and future Dr. Stephanie Maatta-Smith—two of the terrible three. Girls, girls, girls…what have we done? Lord

help this field while we are afoot. Future Dr. Bruce Smith for quietly dealing with the terrible trio. Claudia Montague, the fearless researcher and the endless supporter. Jamie Madden—Melt Banana, thank you! Aimee Reist, the true Librarian, no words will suffice, dear. Dan and Leslee for being the best neighbors.

Thanks to all the Goldstein Library staff who I terrorized in one way or another over the last few years. Thanks to all my past and present students—any time I felt like packing it in I just had to go to class and you picked me right back up again.

I must also acknowledge the wonderful assistance and/or support of the following: Case Western Reserve University Archives, Dr. Jonathan Furner, Alex Justice, Kerstin Michaels, Dr. Michele Besant, Dr. Marcella Genz, Dr. Ron Blazak, Dr. Myke Gluck, Dr. Charles Conaway, Ruth Woo, Dr. Elizabeth Logan, Kandi and David, Kathy and Brian, Paula and Matt, Scott and Lisa, and the many others I'm sure I am forgetting. P.D.B.W., H.V., and D.L.S., thank you.

My enormous family—Mom, Dad, Mary, Beth (first is best, Dear), Francis, Joel, Stacy, Dodie, Marci, and Ron. The nieces and nephews: Heather (dearheart, years from now when you read this, you will understand what part you played in its creation), Michael, Allison, Shannon, Kelsey, Nicholas, Patrick, Ember, Jeremy, Emma-with-the-one-sock, and any others that will surely come along. I love you all very much. To my future husband and children—do hurry up along.

To Dad, especially, what could I possibly say that would even come close to expressing anything?

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## ABSTRACT

An author co-citation analysis was undertaken to determine the relationship between library classification research and information retrieval research during the eighteen year time span of 1952 to 1970. The Classification Research Group (CRG) and the Center for Documentation and Communication Research (CDCR), two of the best known research groups in the two areas in question, provided authors to be used in the analysis. A cluster analysis revealed four distinct clusters among a sample of twenty-one authors. These four clusters were composed of classificationists, automated retrievalists, theorists, and evaluators of information retrieval systems. Multidimensional scaling placed the authors on a two-dimensional map in accordance with proximity values and revealed either intellectual distance or closeness among the authors. A factor analysis resulted in the twenty-one authors loading on four factors that closely resembled the cluster characteristics with some additional author cross-over among the factors. From this trio of multivariate analyses the two dimensions of the map, the X-Y axes, were determined as use of mechanization (Y) and the conceptual view (X). As clusters move left to right a shift in conceptual view from a holistic to an atomistic overview of information organization and retrieval was observed. The use of mechanization decreases as one moves down the vertical axis. Application of three concepts taken from the theory of normative behavior—social types, worldviews, and information behavior—aided in

interpreting the behavioral aspect of the CRG and the CDCR members as revealed by the author co-citation analysis. Future research will include a bibliographic coupling analysis and a complete content analysis of the published works involved. These two analyses will provide further insight into the degree of conceptual similarities and differences between the CRG and the CDCR during the eighteen time period in question.

# CHAPTER 1

# INTRODUCTION

## Problem Statement and Context

C.P. Snow (1959) gave his famous Rede lecture on the "two cultures" with the intent of bringing to light the gulf that existed between two different "polar groups"--literary intellectuals at one pole and physical scientists at the other (p. 3). His examination was of separate disciplines, each with its own separate goals and methodologies that "…had almost ceased to communicate [with each other] at all" (p. 2). While the delineation between these two groups is extreme and easy to distinguish—hard science versus literary studies—it is not as easy to distinguish between specialties within one discipline, especially when considering the intersection and direction of ideas and research within the specialties themselves. This dissertation examines within the field of library and information science two sub-fields: classification research and information retrieval research. Each has become perceived as a separate area by most in the field without much in the way of interrelationship. There has been no systematic analysis of either how or why this perception developed. This study proposes to characterize and analyze the relationship between these two sub-fields by performing parallel case studies of a well-known group of researchers from each sub-field.

**Sub-field 1: Library classification research**

Library classification research is defined as the creation and application of structures of knowledge categories that allow for effective arrangement and retrieval of information entities and illumination of the relationships between those information entities. Traditional library classification research refers to the identification of the structure of the whole of recorded human knowledge with the aim of creating schemes based on that structure in order to organize, show relationships between, and provide access to the base of published information. While more recent work in classification has focused on schemes for specialized areas of knowledge, the focus of research just prior to the mid to late 1950s and early 1960s was on large universal systems—such as the Library of Congress Classification (LCC), Dewey Decimal Classification (DDC), Colon Classification (CC), Bibliographic Classification (BC), and the Universal Decimal Classification (UDC). The work of that period did not emphasize everyday use of the schemes, although such use is an integral part of librarianship. But, behind such schemes—both specialized and universal—is a commitment to seeing knowledge as structured.

**Sub-field 2: Information Retrieval Research (IR)**

Information retrieval research (IR) is defined as the creation and application of computerized techniques for retrieving information entities. These techniques include latent semantic indexing modeling, query language development, and cluster analysis, to name just a few. IR research began in earnest in the 1950s and was well on its way to sophisticated developments in the 1960s. Presently, information retrieval is the focus of

a significant portion of the research conducted in library and information science. Although information retrieval research uses classificatory methods to group both information entities and their contents, it has not traditionally extended the use of classificatory techniques to create the type of schemes used in traditional library classification systems. So, while these two research areas share the common goal of improving methods for the retrieval of information entities, the approaches taken by the two research areas became increasingly distinct in the mid to late 1950s and early 1960s, resulting in the perceived division.

**Perceptions of Sub-fields**

A casual reading of two information retrieval texts reveals the perception of traditional library classification. For instance, in their classic text on information retrieval, Salton and McGill (1983) describe "conventional" library classification as a technique "where the library items are placed on shelves according to their subject content" (p. 215). Library classification is viewed only within the context of the physical placement of the books. Physical location is an important aspect in classification research, but it is not the sole purpose of a classification scheme. The larger issue at stake is the sense of direction the structure gives to the user in terms of trying to find the information the user needs within the realm of knowledge. This is seemingly ignored in information retrieval research and attention is instead focused on examining a document, or collection of documents, for index terms or keywords and classifying those words according to similarities and statistical characteristics on an as-needed basis (Salton and

McGill, 1983). This is contrasted with the additional act of placing the main ideas of a document within a larger overall structure as is the focus of classification research.

In a more recent textbook, the emphasis has shifted towards "modeling, document classification and categorization, systems architecture, user interfaces, data visualization, filtering, languages, etc." (Baeza-Yates & Riberiro-Neto, 1999, p. 2). While their statement displays a broader perspective of the issues involved in information retrieval, the authors still ignore the accomplishments and scope of traditional library classification. This is especially the case with reference to the World Wide Web, which the authors view as a "universal repository of human knowledge and culture" (p. 2). The authors' concern is how users add to and search for information within the Web; further, they suggest that an obstacle in searching this repository is "the absence of a well-defined underlying data model for the Web, which implies that information definition and structure is frequently of low quality" (p.3). This is not a new problem nor one confined to the web environment. Traditional library classification research has sought to fill this need for an underlying model of knowledge since the late 19th century. Indeed, one classification researcher writes, "there are signs that Internet culture is groping, albeit ponderously, in the direction to which 20th century classification thinking has been pointing" (Coates, 1997, p. viii).

For years classification researchers have been stating bluntly and quite succinctly the need for classification techniques and theory in information retrieval research. The need for this combination has been noted over the years by different researchers. In the early 1980s, Ingetraut Dahlberg, an influential figure in the area of classification research, complied a three-volume survey of the "entire post-WWII classification and

indexing literature…as well as the literature on these and their [related] theories, methods, implications, and use" (Dahlberg, 1982, p. iv). The internal classification system Dahlberg developed for the three volumes—called the Classification Literature Classification (CLC)—encompasses not only the literature of traditional classification but also some of the literature emanating from information retrieval research.

Joan Mitchell, current editor of the Dewey Decimal Classification (DDC), views classification as being extremely relevant to the full-text online future because "it will be harder and harder to effect meaningful retrieval without devices that supply context and relationships …classification, whether overtly through notation or covertly as an underlying navigational aid, provides a powerful tool for both the intrepid and the timid information web explorer in library catalogs and beyond in the global information network" (Mitchell, 1995, p. 45). Mitchell's comments about the Web echo the stance of many classification researchers on the usefulness of classification for electronic information retrieval in general from as early as the late 1950's.

Throughout the 1980s and 1990s there was a resurgence in interest in the traditional classification structures with regards to online information systems. The integration and upgrading of traditional classificatory techniques was the object of numerous exploratory studies. A good example of this is what Dahlberg calls "the Karen Markey Shock of 1984." Karen Markey and Anh Demeyer, in a study sponsored by the Online Computer Library Center (OCLC), conducted an evaluation of the integration of the DDC schedules and index into an online catalog and found that it "provides new strategies for subject searching and browsing that are not possible through the alphabetical approaches of subject headings and/or keywords presently supported by

online catalogs" (Markey and Demeyer, 1986, p. xliii). Dahlberg (1995) believed this study brought classification "back into the minds of doubtful librarians and all those who thought they would not need it any longer" (p. 23).

Currently, there is evidence of even more interaction between the participants and scholars involved in the two research areas, especially when dealing with the ever increasing challenges presented by the Internet and the World Wide Web. If one examines Morville and Rosenberg's recent book *Information Architecture for the World Wide Web* (1998), there is an obvious reintroduction of more traditional classificatory structures in the developing area of information architecture. For example, the authors devote a significant portion of their chapter on organizing websites and intranets to a discussion on the type of organizational structures and schemes that have been traditionally confined to classification literature.

The interest in taxonomies and ontologies has been very apparent in Web search indexes such as Yahoo! and the Open Directory Project, both of which employ a pseudo facet-analysis approach to indexing websites. Facet analysis is unique to the field of library classificatory method and was first introduced by the classificationist S.R. Ranganathan in the mid-20$^{th}$ century. This analytical-synthetic process consists of breaking down a subject into its constituent elements with respect to a parent, or main, class. For instance, a website on the Der Blaue Reiter movement in art is given the heading "Arts>Art History>Period and Movements> Expressionism> Der Blaue Reiter."[1]

The American Society of Information Science & Technology (ASIST) 2000 Classification Research special interest group (SIG/CR) workshops demonstrate a high level of interaction with research between classification and information retrieval

research, especially in research conducted on classification in user interfaces, classified displays of search results on the Web, automatic indexing using categories and automatic concept hierarchies development (Soergel, 2001). This increased interaction is largely the result of the influence of the International Society for Knowledge Organization (ISKO) of which a numerous portion of the SIG/CR participants are also members. This organization was established in 1989 by Ingetraut Dahlberg in an effort to keep the knowledge classification tradition alive by bringing together researchers from the field of library and information science, philosophy, linguistics, computer science, as well as such special domains as medical informatics (ISKO website, 2001).

But at the same time the general attitude among many information retrieval researchers is one of casual ignorance or of hesitation in viewing classification research as beneficial to their work. Matt Koll (1999), in his comments on the information retrieval track papers at the 1999 Annual American Society for Information Science (ASIS) conference, began by describing himself as "…a search guy, a full-text guy. I've never been a classification person." He defined classification as *information losing* and went on to say that

> …when you put something into a category you tend to not maintain all the other pieces of information that ever went there. You tend to put it somewhere and sort of erase, or minimize, the other traces of information and I just think we need to be really, really, careful about that, particularly one of the trends…is the growing integration of classifications and full-text search and the emergence of *ontologies* and *taxonomies* as the buzzwords of 1999 and 2000. We really do have an opportunity to make

> classification information *producing* which, when done well, it is (Koll, 1999).[2]

It is an accepted fact that the categorization of knowledge does lose information because the act of categorization involves selecting some aspects of an information entity and not others--hence, some information will be lost. But, again, Koll's narrow comments reflect the steadfast opinion that the emphasis of library classification is simply on its functional process—that of placement, or pigeonholing, of knowledge—and not the deliberate conceptual structuring of knowledge. This perception of traditional library classification can be traced back as early as the 1950s and 1960s when work in IR began in earnest.

It is obvious there are some fundamentally different points of view at work and that these differences have led to a division between library classification research and information retrieval research. Birger Hjørland (1999) suggests a difference in conceptual views that has served to push ideas and research into different, often conflicting, directions. He writes,

> …classification research and IR are two scientific traditions building on different epistemological views or ideals. Each of these traditions has something to contribute to IS, and it is urgent that this is recognized, that the traditions are brought together and do not continue to exist as separate traditions (p. 475).

The epistemological view he describes is a difference between *holistic* and *atomistic* views. IR research focuses on the atomistic view of documents—the breaking of a document into "very specific concepts and words”—and also with evaluating information

systems by measuring the recall and precision of information retrieved from the systems via this atomism of words and concepts (Hjørland, 1999, p. 475). As defined previously, traditional classification research generally takes a more holistic and conceptually-oriented approach. It deals with the overall structure of the knowledge as applied to a collection of documents. Hjørland shares the opinion with others that a combination of representations would be more beneficial to LIS in general. Similarly, Jack Mills, writing on classification in information retrieval, remarked:

> It is often assumed that the organization of knowledge in libraries is insufficiently important to warrant continued research. I think this is a very short-sighted view. A truly comprehensive, flexible and logically structured map of modern knowledge, designed expressly to serve its central functions, is surely the least the library and information profession deserves now (Mills, 1998).

Mills, along with Brian Vickery, Eric Coates, and Douglas Foskett, was one of the original members of the Classification Research Group (CRG), a British-based group dedicated to developing classification systems for the specific purpose of scientific information retrieval. The members assembled themselves in the early 1950s, but it was in the mid to late 1950s through the 1960s that their most influential research was conducted and presented. At the same time in the United States, the Center for Documentation and Communication Research (CDCR) at Case Western Reserve University[3], developed by Jesse Shera and operated in its early years by Allen Kent and James Perry, was coming into its own by providing a much needed stabilizing and marketable force in the quest to control the steady publication of scientific information.

Both groups had international recognition and both were represented at the renowned International Conference on Scientific Information (ICSI) held in Washington D.C. in 1958. This conference was conceived "to bring together on an international level scientists and information specialists for discussion of current research progress and problems concerned primarily with the storage and retrieval of scientific information" (Atwood, 1959, preface). It has been described as an "exhilarating experience" that served as a "status report that...pinpointed areas for further research" (Farkas-Conn, 1990, p.189). Ultimately, the organizers hoped that the conference would "stimulate further research and cooperation among those who are attempting to cope with the problems involved in making scientific information easily and rapidly available" (Atwood, 1959, preface).

## Plan of Study

The CRG and the CDCR represent the most recognized and organized groups within each of the two areas of classification research and information retrieval research from 1952 to approximately 1970. This study will analyze the work of these two groups for the purpose of characterizing the relationship between classification research and information retrieval research during this time period.

Furthermore, this study will investigate the two groups' perceptions of general library classification schemes such as the Library of Congress Classification, Dewey Decimal Classification, and the Universal Decimal Classification, in an effort to illustrate the use of practical versus theoretical IR classificatory mechanisms. The CRG wanted to change these schemes from within--to modify them for better use in information retrieval.

IR researchers of the time, like those at the CDCR, equated library classification research with these systems and not the underlying theories and techniques that were used to develop them. As such, they chose to move away from these general systems altogether and devote more energy to machine-based document retrieval techniques.

This investigation will involve several stages. First, a citation analysis will identify the clusters of authors via the level of co-citation between authors(number of times two or more authors are cited together). Second, theses clusters will be characterized in order to determine where the work of the two groups either came together or diverged on a conceptual level as pertains to both the techniques and systems used to organize both scientific and non-scientific literature, as well as how the conceptual similarities and differences may have changed during the eighteen year time period in question.

The clusters of authors that result from the co-citation analysis will also be examined through the lens of the *theory of normative behavior*. This theory seeks to explain "…the ways in which people deal with information in the contexts of their small worlds" (Burnett, Besant, & Chatman, 2001, p. 536). By *normative* is meant "a typical and natural way of relating to a norm or standard of conduct" and it is in this way that it will be used in this study (p. 536). The concept of *small worlds* is defined as "…a specific context that serves a particular population to permit its members to conduct their business in a routine, expected manner" (p. 536). It is hoped this theory will be helpful in understanding the differences in the two groups by shedding light on where and how the mutual opinions and worldviews of these two groups either overlapped or diverged as evidenced in their published work.

There are four concepts used within the theory of normative behavior: worldviews, social types, social norms, and information behavior. Each one will be examined to determine if it is a viable tool for explaining the relationship between the two groups.

*Worldview* is a "collective perception held in common by members of a social world regarding those things that are deemed important or trivial" (Burnett, Besant, & Chatman, 2001, p.537). For instance, classification is closely allied with library-based document retrieval, while IR research emphasized machine-based document retrieval. Late 19th century and early 20th century library classification research was greatly influenced by the idea of a natural order, or structure, of knowledge and this in turn related to documents in that the goal was to try to 'fit', or classify, documents into that order or structure.[4] Machine-based document retrieval focused on techniques for sorting documents and while this *may* have been related to an overall knowledge structure, those creating these document retrieval systems were not really interested in that overall structure. Stated another way, IR research narrowed the notion of document retrieval from the intersection of a person's question with a collection of documents organized in a structure of knowledge to an intersection of a person's question with a collection of documents that are not organized within any structure of knowledge. In this view, IR seemed to conclude that a structure of knowledge would not be necessary in order to retrieve documents that would answer a question. In essence, while traditional classification researchers would ask *Where in the structure (e.g., the DDC or UDC) are the documents with the answer?* the information retrieval researchers would simply say *No, in what documents will the answer be found?* By an examination of the citation

patterns produced by the citation analysis these two differing worldviews can be more fully understood and perhaps verified.

A second concept in the theory of normative behavior is that of *social types* which "pertains to the classification of a person or persons, and are the absolute definitions given to individuals within a social world" (Burnett, Besant, & Chatman, 2001, p. 537). This study will attempt in some ways to define the "social types" of the members of both the CRG and the CDCR. To accomplish this examination, the post-WWII movement in the library and information science field towards automation provides some important clues that must be acknowledged. To the information retrieval researcher classification was a tool, but the focus was on the information within the documents themselves. The increase in automated techniques after WWII merged in some ways and conflicted in other ways, with the traditional library practices.[5] The movement toward automation can also be attributed greatly to the post-WWII interest of scientists and engineers in organizing their own literature and their endeavors to produce machine-retrieval techniques. Indeed, many scientists and engineers at this time began new careers in the library and information science field. However, Farkas-Conn writes that

> …given the wartime and postwar involvement of scientists and engineers in information work, it is astonishing that coding and the terms used for retrieval--by whatever name--and their use for different kinds of collections were not approached more systematically… scientists and engineers may have understood better than librarians the underlying structure of science and technology….in information work,

however, they did not use proven approaches of their disciplines (1990, 135-136).

The LIS field in the 1950s and 60s has often been characterized in terms of the division between "librarians and documentalists," where some librarians were perceived as not wanting to update their procedures and some documentalists were convinced that automated methods such as punch card technology and microfilming were the solutions to the problem and that moving away from traditional library approaches was the first step in putting these new solutions to work (e.g., Bowles, 1999; Farkas-Conn, 1990). These elements—increases in scientific specialization and scientific publication and in the use of automated techniques—resulted in a change in the practices and as a further result brought about a wish "to gain a better understanding of the fundamental problems" that were appearing in LIS at the time (Farkas-Conn, 1990, 145). Through an examination of the *social types* this study will ask if the conflict between librarians and documentalists, as discussed by both Bowles and Farkas-Conn, can also be applied to the members of the two groups. If so, can we characterize the social types in the CRG simply as "librarians" and the social types in the CDCR simply as "documentalists"? One of the difficulties with this characterization may be the fact that memberships within both the CRG and the CDCR were peppered with scientists and practicing librarians. What role or type did the members of either group see themselves as occupying?

Within any given small world there are set boundaries as dictated by s*ocial norms.* These norms help to hold a small world together and "give individuals a way to gauge what is 'normal' in a specific context at a specific time" (Burnett, Besant, & Chatman, 2001, p. 537). For instance, Burnett and Besant, studying virtual communities

and feminist booksellers respectively, found the social norms of each group centered around such things as "unspoken guidelines about verbal structure of texts" (virtual communities) or "agreeing to disagree" when defining themselves (feminist booksellers) (p. 542-543). By a similar argument, could the social norms of the CRG and the CDCR be defined by the citing behavior as revealed by the citation analysis? Were there "boundaries" where citing certain author's works were concerned and did the members of either group feel inclined or disinclined to cross those boundaries?

Closely related to the guidelines for citing behavior is the actual behavior of who cites whom and how often. This can be viewed as the *information behavior* of the small worlds of the CRG and the CDCR. This concept is defined as "…a state in which one may or may not act on available or offered information" (Burnett, Besant, & Chatman, 2001, p. 537). For instance, in the 1950s and 60s the CRG were also interested in the use of automated methods coupled with the fundamental classification principles and were, in a sense, attempting to move away from the idea of a "natural order" in the classification of knowledge—in short, to break with the accepted tradition or worldview—of their area of study. These accepted worldviews had a longer history than those accepted in IR research at the time. So, at the same time library classification researchers were trying to break with the older traditions they were also trying to merge with the new work being done in IR. Do the citing behaviors of the CRG reflect the need to move away from the older ideals of classification research, i.e., were they citing more IR authors or were they remaining within their own small world of accepted classification writings? In addition, was this "move" acknowledged by those outside of their small world, i.e., who was citing them and did they cite the work of the CDCR at the same time? Similarly, the

information behavior of the CDCR can be determined along the same lines. Was there a degree of overlapping citing behavior and if so how can this be used to ascertain the relationship between classification research and information retrieval research?

I am convinced that by examining and comparing the work of the CRG and the CDCR with the methodologies discussed above, we may begin to view and understand this relationship at a more granular level. In essence, this project is an attempt to understand the relationships the *small worlds* of traditional classification researchers and information retrieval researchers through the use of citation analysis and the application of the *theory of normative behavior*. Did the small worlds of each group overlap for a time and then began to shift away from each other? Alternatively, had they been completely separate from the start and perhaps paralleled each other in terms of research and progress in their respective goals of organizing and providing access to information?

## Background of the problem

The late 1950s were highly active where publication and consumption of scientific information was concerned. Vannevar Bush described this as an "information explosion" in his seminal 1945 article "As We May Think." Sputnik was launched in 1957 and spurred the United States to improve upon and surpass the Soviet's technical and scientific achievements. For the established field of library science and the growing field of information science the challenge was to develop systems and methods for keeping up with the pace of the publications. Helen Brownson, who at the time was program director at the National Science Foundation's Office of Scientific Information and later a co-founder of *Annual Review of Information Science and Technology*

(ARIST), recalled that this time period during which the Soviets were demonstrating technical achievements

> …contributed to a change of climate surrounding information science and also our NSF efforts to develop programs aimed at facilitating access to scientific literature. 1958 became the busiest year in my recollection. The President and his Science Advisory Committee, headed by Dr. James R. Killian, took an interest in the science information problem, The Committee named a special subcommittee, headed by Dr. W. O. Baker, Vice-President for Research of Bell Telephone Laboratories, to look at the problem of improving access to scientific literature (Brownson, 2000).

Within this environment there was not only did the issue of the growing division between librarians and documentalists exist, but also a difference between American efforts and European efforts, this despite the air of global cooperation as evidence, by the International Conferences on Scientific Information (ICSI) held in 1958. Jack Mills, in a letter to this researcher, speculates that differences in classification and information retrieval research may be rooted in the "cultural differences between Europe and America in their attitudes to the connections between conceptual analysis and unruly language as instruments in IR" (Mills, personal communication, 2001). Eric Coates, similarly suggests that the US and the CDCR located there were "decidedly more oriented towards the mechanization issue than was CRG at the time… CRG was set-up to investigate the subject retrieval process in the (for UK) pre-mechanization environment" (Coates, personal communication, 2001).

**Classification Research Group**

The Classification Research Group officially formed in 1952 after it was suggested at the 1948 Royal Society Scientific Conference that "a study of classification should be made" (Foskett, 1962, p. 127). This rather simple suggestion led to the formation of a group of professional colleagues that Foskett described as "…a typical British affair, with no resources beyond the native wit of its members, no allegiance to any existing system of classification, no fixed target, no recognition by the British Government (naturally), and at first only an amused tolerance from the library profession" (p. 127). The group's first action was to agree that the starting point for their work would be the rejection of all existing classification schemes because they were "unsatisfactory." This dissatisfaction has stemmed from the view that theses schemes were unable to support the *translation* of the concepts found in documents into concepts and term choices that could be expected of the language of the queries put forth by researchers. In other words, matching the user's query with the correct document could not be accomplished adequately enough with the existing classification schemes that were used as the "master stereotype" for organizing the documents (p. 128). The group then turned its focus to reviewing the "basic principles of bibliographic classification unhampered by allegiance to any particular published scheme" (CRG, 1955, p. 262). At the end of their review it was decided that a faceted classification scheme should be the basis of all methods of information retrieval. In essence, a faceted classification scheme allows for "a given genus to be subdivided in more than one way, to give several sets of subclasses, each of which is a homogeneous group of collateral species" (p. 267). As Calvin Mooers described it, it is an analytical tool that "allows you to peel the onion of an

idea" (Mooers, as quoted in Vickery, 1966, p.14). The CRG believed this level of analysis of concepts in documents would better serve retrieval of documents in information systems rather than the traditional top-down method of assigning subjects to main classes and sub-classes in traditionally accepted hierarchical schemes. To this end, members of the group—both central and peripheral—began investigations into both the inadequacies of current general classification schemes and also detailed examinations of special subject fields with the idea of "further developing principles [of classification] and of constructing satisfactory [classification] schedules" (CRG, 1955, p. 268).

The CRG was productive through most of the 1960's, with both group publications and individual member publications. Bulletins describing the work of its members were published in the *Journal of Documentation* four times during that decade, but in the 1970's and 1980's the bulletins became much more infrequent.[6] In addition to developing faceted classification methods, they also investigated using the theory of integrated levels to build a new general classification scheme. Quite simply, integrative levels are regarded as “simple linear series of complexities increasing upward in the scale” (Feibleman, 1985, p. 142). This theory stressed "that the world of entities evolves from the simple towards the complex by an accumulation of properties or influences from the environment" (Foskett, 1978, p.204). Subject analysis allows for the identification of the important concepts—subjects—in any given document. The theory of integrative levels proposes that the level of detail of analysis should not go beyond "the level at which the unit [represented by a term] acts as a whole and in a particular way for a particular purpose" (Foskett, 1961, p.148). In their search for the principles of classification, this theory provided some interesting opportunities as well as some

interesting problems, such as the sequencing of classes in a classification scheme (Spiteri, 1995).

In 2000-2001 the aim of the group is focused on developing the Bliss Classification system and they meet regularly to discuss the development of the remaining schedules of that system which is not as widely used as they would wish.[7] They have also turned an eye towards the future of classification research because "it has always been a primary objective of the CRG that its work is founded in practical classification, and in the application of theory to specific situations; it may be that in the 21st century the understanding of the situations to which the theory is appropriate needs re-examining and restating" (McIlwaine and Broughton, 2000, p. 199).

**Center for Communication and Documentation Research**

The Center for Documentation and Communication Research (CDCR) was founded in 1955 at the School of Library Science at Case Western Reserve University[8] (SLS/CWRU) and quickly established its leadership in the realm of punched card technology that allowed for the quick searching of abstracted scientific and industrial literature. Jesse Shera, then dean at SLS/CWRU, established the Center as a department within the School and appointed James Perry and Allen Kent as director and associate director, respectively (Shera, 1971, p. 223). Although members of a library school faculty, Perry and Kent did not have traditional library backgrounds or training. Perhaps it was this non-conventionality that helped to propel the CDCR's reputation for information storage and retrieval during a time when traditional library practices were being unfavorably viewed by those within the scientific and technical industries. These

industries produced vast amounts of research and publications and as such needed help when it came to literature control and information access needs. Kent and Perry each had training in chemical research and shared a strong interest in the problem of controlling the growing amounts of scientific literature (Farkas-Conn, 1990, p.140). Perry had been at MIT during the 1940s and 50s and it was there he was exposed to and took part in the development of machine translation of documents (p. 141). In 1945, Kent took part in the organization of rocket development documents taken from the Germans at Peenemunde and became intrigued with the work of the library profession. Upon his return to the U.S. from Germany he began work as a research chemist and chemical engineer, but in the early 1950s he went to work at Interscience Publishers because the opportunity to return to documentation work and its problems "became irresistible" (Kent, 1961, p. 772).

The CDCR's goals were to conduct research into the "newly developing non-conventional methods of information storage and retrieval and to enrich the educational program of the school by introducing new courses and seminars in this area to provide opportunities for students to become engaged in the research and operational program of the center itself" (Shera, 1971, p. 223). Between 1955 and the Center's demise in the early 1970's it had interactions with and funding from a variety of commercial and industrial corporations, and the U.S. government.[9] IBM, General Electric, Eastman Kodak, Shell, Goodyear Aerospace, and various pharmaceutical and chemical companies all made visits to the Center (Bowles, 1999). CDCR also developed and worked on special documentation projects for the American Society of Metals and for the Communicable Disease Center (CDC), now the Center for Disease Control, and the

American Diabetes Association as metallurgical and biomedical information were "areas in which the CDCR excelled" (Bowles, 1999, p. 260).

Perhaps the most significant development to come from the work of the CDCR was its procedure of semantic factoring and the resulting semantic code system developed by Perry, Kent and Madeline Henderson, who worked with Perry and Kent at Battelle Memorial Institute in Columbus, Ohio, before they joined the CDCR in the mid-1950s. Machine searching of literature relied on subject headings, or descriptors. These terms could then be coded. However, the codes themselves "could, but did not have to, reflect the meaning of the term they represented" (Farkas-Conn, 1990, p.140). Perry believed they did and, because he saw information as "polydimensional," he wanted the coding system to reflect the complexity of the vocabulary of subject terms (Farkas-Conn, 1990, p.140). In essence, what was being developed was a controlled vocabulary with "indications of the conceptual relationships between terms and of synonyms and near synonyms and giving also a grouping or classification of those terms" (p. 141). The controlled vocabulary thesauri that resulted from this work assisted in the machine scanning of indexes and abstracts via punchcard technology. It was this technique that allowed the CDCR to attract the attention of the commercial and industrial organizations who were losing money because of the "ineffective utilization of recorded knowledge" (Kent, 1961, p. 773).

**Pigeonholes and Punchcards**

The CRG and the CDCR shared the belief in the inadequacy of general library classification schemes as they were being used at the time and in their application for information retrieval purposes. The CRG saw problems "arising out of the inability of the well-known schemes to cater for the complexities of modern knowledge and the demands of modern library services" (Foskett, 1963, p. 11). They were especially unsatisfactory for specialized fields of knowledge. The CDCR labeled these general classification tools "pigeonhole classification" because they operated under the convention that documents with like characteristics will be grouped together under broad or generic headings, i.e., placed into pigeonholes (Perry & Kent, 1958, p. 34). Kent and Perry saw this as a disadvantage because of the practical limitations: "[I]t is often not possible to predict which combinations of characteristics will provide the most useful grouping of the documents involved" (p. 35). They sought to escape from the pre-coordination aspect of traditional classification "by applying various mechanical and electrical devices ranging from hand sorted punched cards to electronic computers." (p. 37).

By the mid-1960's, these *small worlds* appear to have begun diverging. The CRG, despite their criticism of traditional schemes, did not abandon them altogether but rather wanted to adapt the schemes from within. The CDCR, then leaders in information retrieval research, basically saw no use for them and did not include them in their work at all.

## Research Questions

This research will investigate the citing of the published works of the CRG and the CDCR from 1952 to 1970 in an effort to understand how these works either overlapped or diverged with respect to how the authors sought to provide better organization and retrieval of information and in their use or non-use of general classification schemes. The main goal is to characterize this relationship by answering the following questions:

1. What was the initial relationship between the Classification Research Group and the Center for Documentation and Communication Research and did this relationship change between 1952 and 1970?
2. What are the conceptual similarities and differences of these two groups as evidenced in their published artifacts from 1952 to 1970?
3. How can we characterize the relationship between the small worlds of the two groups using the theory of normative behavior and its concepts of worldviews, social norms, social types, and information behavior?

## Assumptions

There are two main assumptions-- (1) the relationship between classification researchers and information retrieval is such that IR researchers and classification researchers presently conduct their research, for the most part, independently of each other. (2) The Classification Research Group and the Center for Documentation and Communication Research are good representatives of the two areas during the time period being studied.

## Delimitations and Limitations

This is not a comprehensive historical study of the Classification Research Group (CRG) and the Center for Documentation and Communication Research (CDCR), nor is it a study of the current state of IR and classification research. The case study approach allows for in-depth examination of these two groups that would not be possible in a

broader survey, but may limit the generalizability of the conclusions to other research being conducted in this field at the same time.

The use of multiple methods (a citation analysis involving three multivariate analyses, and the application of the normative behavior theory) will provide cross-verification of the conclusions. The time period of 1952 and to 1970 will be strictly adhered to in the citation analysis. Thus, the projection of influence beyond the time period will not be determined, nor will the researcher be able to unambiguously state that all published works by all scholars involved in these two groups was examined.

Potential weaknesses of the study may be uncovered if the citation analysis shows little or no connection between the literature of the CRG and the CDCR. The literature review may not adequately reflect the conceptual similarities and differences being sought or it may disregard the context that produced the text in the first place. Every attempt will be made to make adjustments for these problems as the study develops.

Lastly, this investigation attempts to identify more precisely the relationship between these two areas of research—not to reconcile the two areas to any degree if a division in their work is found.

## Significance of the Study

"If a field does not document its past, it will lack a history and have a diminished sense of identity" (Buckland and Hahn, 1998, 1). Based on the reasons identified in the preceding pages, the significance of this study lies in its attempt to better understand why two areas of research in the field of library and information science diverged when the ultimate goals of each—effective organization and retrieval of knowledge—are so clearly related. The significance also lies in the fact that general classification schemes such as

the Library of Congress and the Dewey Decimal classification schemes are still used today on a large scale despite the persistent criticisms by both the CRG and the CDCR that these systems were inadequate for effective organization and retrieval of information entities.[10] It is important to ask and find out if a bifurcation of the research of these two areas may have contributed to circumventing the problem of organizing and accessing so much literature, rather than to solving it. This is especially important in the current web-based environment which has so challenged current researchers and practitioners with its seemingly never-ending production of information, much like the post-WWII period.

In 1997 the originating members of the CRG met to remember the International Study conference on Classification for Information Retrieval, which was held in 1957 in Dorking, England. This conference has long been remembered by CRG members as being the most influential to the group's work. A remark by Cyril Cleverdon adequately justifies the significance of this study:

> At the time of Dorking, the use of computers in information retrieval was only an idea and it was impossible to foresee the increase in computer power which would permit the analysis of full text using complex statistical techniques for matching queries and documents. However, in spite of thirty years development along these lines, the gain in performance, as measured by recall and precision, appears to be slight. Perhaps a bringing together of the logic of classification and the

power of modern computers could be a way forward (Cleverdon, 1997, vi).

## Endnotes

[1] An actual faceted classification scheme involves a more complicated construction and involves a complex classification notation.
[2] Koll did not elaborate on or define what he meant by his last comment of "which, when done well, it is."
[3] Western Reserve University and Case Institute of Technology united to form Case Western Reserve University in 1967. While it was not merged at the time the CDCR was formed, this study will refer to it as CWRU, for the sake of simplicity.
[4] This is evident, at least, in the work of the influential classification thinkers of the early 20th century -- Richardson, Sayers, Bliss, and, to some extent, Ranganathan, although his work diverged in many ways from the other three. The work of all four of these men had an enormous influence on the CRG.
[5] Chapter 7 in Farkas-Conn's book "From Documentation to Information Science" gives a detailed analysis of the events of this time and their effect on the practices and concepts of librarians and documentalists.
[6] Bulletins 1-3 are unavailable. Bulletins 4-9, were published respectively: 1958, 1959, 1961, 1962, 1964, 1968. Bulletin 10 was published in 1973, 11 in 1978, and 12 in 1985.
[7] The Library Association (LA) used the library and information science scheme in it Library Association Library until 1993.
[8] See endnote 3.
[9] The CDCR was absorbed by the library school at CWRU and other components of the university.
[10] There have been continual improvements to these systems, i.e., they are not the same as 50 years ago.

# CHAPTER 2

# LITERATURE REVIEW

## Introduction

The review of the literature surrounding this project will involve two types of writings: 1) primary literature, that which was produced by members of the CRG and CDCR during the period in question; and 2) secondary literature, that which comments on the activities of these two groups. This chapter will examine the secondary literature about the Classification Research Group (CRG) and the Center for Documentation and Communication Research (CDCR) as subjects of discussion. This will serve to construct the informing context surrounding the questions that guide this study.

## Historical literature on the CRG and the CDCR

Historical literature written about the two groups has been examined in order to identify the goals and accomplishments of the groups and how those changed over time, the influence of each group on the field of library and information science (LIS), and their views of other methods of information organization and retrieval, particularly those methods employed in America and England. As stated in the introduction both the Classification Research Group (CRG) and the Center for Documentation and Communication Research (CDCR) were formed in the early to mid-1950's in England and the United States respectively. As groups, or institutions, there have been only a

handful of writings published about them. In addition, a majority of what has been written regarding their work has been done so by the members themselves.

**Classification Research Group**

As a group, the CRG has quietly worked in classification research for nearly fifty years. During that time, its members have tackled from many angles the question of creating a universal general classification system for use in the information organization and retrieval systems used to store humankind's knowledge. They began by first declaring they would "review the basic principles of bibliographic classification, unhampered by allegiance to any particular published scheme" (Vickery, 1953, p.187). In effect, they rejected all existing classification schemes—most specifically, the Dewey Decimal Classification (DDC), the Library of Congress Classification (LCC), and Universal Classification System (UDC), which were the major systems used in libraries. This was not a rejection of organizing knowledge via classificatory techniques. They felt "that the systematic organizing of documents by subject could play a major part in developing successful information systems" (Foskett, 1962, p. 27). However, the existing systems were deemed inadequate for a number of reasons:

> …they give insufficient detail; revision and extension are too slow; the placing of subjects varies widely from one scheme to the next; many classes are a heterogeneous jumble of tenuously related terms; subjects occur in more than one class; phenomena whose relations with existing knowledge are uncertain cannot be classed, nor can completely new classes be accommodated; relationships between subjects are distorted; and so on (Vickery, 1953, p.188).

In many respects Vickery's summary of problems found in contemporary general library classification schemes was right on the mark. A brief summary of the status of each is informative. Melvil Dewey's *Decimal Classification* had been in existence for 75 years when its 15th edition was published in 1951, but at that point the issue of who its target audience really was hadn't been resolved. Some had fought over many years for the idea that the primary users of the system were general libraries of modest size that did not need great detail in the system and that benefited from the system's long-standing policy of "integrity of numbers." The latter meant that once a notational position for a subject was established, that position would ordinarily not be changed even though the advance of knowledge indicated a better position for the subject elsewhere in the system. This policy was advantageous to libraries who used the system primarily for the shelf arrangement of books because it saved them from the tedious task of changing numbers on books and in the card catalogs. The downside to this policy was that the system became increasingly out-of-date in its subject collocation. When the latter was combined with the system's use of out-of-date terminology, the result was a system very unsuitable for fast growing areas of science and technology. The opposite position, generally taken by large and growing academic libraries and by science specialists, was that the system simply had to be updated and kept up-to-date no matter what the effect. The 15th edition attempted to serve both extremes. It extensively updated terminology and re-collocated subject areas, but it also drastically reduced the system in size and detail. This edition almost caused the system to fail and was the initial cause for many libraries to shift to the Library of Congress Classification (LCC) throughout the 1950s and 1960s. The DDC's details were restored in great part by the 16th edition in 1958 but it was not until the 17th

edition that steps were taken to change it systematically in an ongoing way. Miksa (1998) writes that by then it was too late to have gained the interests of those in information retrieval.

The Universal Decimal Classification (UDC), in contrast to the DDC, was developed especially for scholarly and technical use. It was begun in 1895 by Paul Otlet and Henri LaFontaine of the Institut Internationale de Bibliographie (IIB) in Brussels, Belgium, at first simply as an enhanced version of the DDC's 5th edition. Soon after its beginning, however, sophisticated faceting devices for indexing scientific and scholarly literature at a very detailed level were added to it and the system continued in this vein until World War I slowed its progress. The IIB was restructured in the mid-1920s and over the next two decades refocused more centrally on science and technology, changing its name in the 1930s to the Federacion Internationale de Documentation in recognition of its desire to provide indexing for science and technology. The UDC itself was produced in a first edition in French in 1905, and a second also in French between 1927 and 1933. Third full editions were begun in German and English beginning in 1933 and 1936 respectively, but World War II greatly slowed their progress. The German full edition was completed in the post war years, but an abridged English language edition was only completed in 1961. From the 1930s to the 1960s the UDC was developed in a very decentralized way among volunteer agencies. This approach yielded many strong revisions, but it proved to be very slow and it contributed to the system developing in a somewhat hit or miss fashion. Also, the system did not really shake off its DDC foundations until after the 1960s. So, while it tried very hard to be a worthy alternative to

the DDC especially for science and technology, it did not give that appearance during the period of the present study. (Foskett, A.C., 1973; Lloyd, 1976)

The Library of Congress Classification (LCC) differed from the other two schemes in having been developed for a particular library—the Library of the United States Congress—with little reference to the needs of the other agencies. Because the library's collections were very large, the LCC tended to be very detailed right from the start, at least in comparison to other systems. As a result, other large libraries, especially those in major universities tended to find it more satisfactory, especially in the social sciences and history, which were the system's initial areas of greatest development. Unfortunately, there was a downside to the scheme. First, after a beginning period of development up to about 1915, classification at the library languished and many more decades were eventually required to complete the scheme. Second, not all areas of the scheme were given equal attention, and it tended to grow erratically. For example, science and technology sections of the system, those that would have been of greatest relevance to the burgeoning areas of science and technology during and after World War II really did not witness great expansion until after 1960. Third, for many years, obtaining up-to-date schedules was not always easy. Lastly, and probably most importantly, although the system contained a great deal of detail, much of it was not based on subject categories, but rather on breaking up large accumulations of materials in a subject area by aspects other than subjects—for example, by the forms of materials, by their textual languages, and by their dates of publications. Further, where subjects were the categories in question, arrangements were often little more than pragmatic alphabetic lists of topics instead of attempts to determine more logical bases of grouping. As a

result of these various factors, the LCC was ill-equipped to handle the kinds of subject access needs that information organizers in the 1950s and 1960s thought necessary (Miksa, 1984).

Given the foregoing account of the state of contemporary general classification systems, it is obvious that a new approach was needed. The CRG sought discussion and experimentation from its members for the undertaking a new approach to classification. Naturally, in rejecting the general systems, the idea of specialized classification systems came into focus. This was quite in line for the time because of the environment surrounding library and information science. As stated in the introduction, the increase in scientific specialties, and consequently scientific publications, sent librarians and documentalists scrambling to produce more powerful and efficient methods of analyzing, organizing and accessing the literature. The CRG was "deeply conscious of the impending revolution in information technology" (Gilchrist, 1997, p. iv). Within this context, the use of computerized techniques was inevitable. The question was simply how they would be used. Jack Mills (1997), reflecting on the work of the CRG, wrote "that the bedrock of any information system is the structure of knowledge seems unarguable; and classification is the key to its comprehension and its practical management" (p. xi). With the rejection of the existing schemes the discussion turned toward "first principles and the foundations of classification methodology, in particular the need for flexibility in combining concepts: pre-coordinate and post-coordination, facet and relational analysis, and the use of roles and links" (Foskett, 1971, p. 141). They were essentially looking for a way to begin "from the opposite end of the road…instead of breaking down the universe of knowledge into basic classes, and analyzing these to

arrive at the individual concept, would it be possible to start by organizing the concepts themselves?" (Austin, 1969, p.151).

The idea of a "universe of knowledge" was, and still is in many ways, the core of traditional library classification research and the general schemes that were produced out of that research. Basically, it assumes that "the universe of knowledge is in fact a universe of classes [disciplines], which we presume in some strange way, have a prior existence accepted by all" (Austin, 1969, p.151). The construction of a general classification system had always been based on dividing the main classes in "successive stages until, in theory at any rate, every conceivable concept can be located" (p. 151). For the reasons stated above, the CRG found fault with this approach and it was decided in the beginning years of the group that "there should be no more of these assumptions" (p. 151). They turned instead to facet, or analytico-synthetic, analysis as first put forth by S.R. Ranganathan in his Colon Classification system in the 1920s. The group "realized that Ranganathan's ideas and theories had more to offer than those of most other writers, and decided to adopt at least some of the Colon terminology and method" (Foskett, 1962, p. 128). Facet analysis hinges on the "subject description of documents" in that subjects are

> …divided into groups known as "facets", and within each facet they may be arranged hierarchically. The facets are listed in the schedule in a prescribed order which is usually the order in which terms are to be combined to form compound subjects. By means of this combination order the relations between terms are displayed (Vickery, 1960, p. 9).

The CRG did not accept Ranganathan's idea of there being five "fundamental" categories of a subject--Personality, Matter, Energy, Space, Time-- but instead decided that "although one examines a subject in the first place in order to determine the categories into which it terms may be arranged, one should not try to force them into facets based on an artificial division, which appear in all schemes" (Foskett, 1962, 129). Essentially, they did not believe in *a priori* categories, but rather would derive the categories of a subject from an examination of the literature itself. Frohmann (1983) examined this position with an eye toward semantics and concluded that the CRG, particularly in this early period of their work, set the stage for the use of empirically derived *a posteriori* semantics for use in their classification systems--meaning that "semantic relations between terms are based upon human activities rather than on *a priori* relations between meanings, conceived as determinants of and prior to those activities (p. 14). Much of the early research coming from the group dealt with finding "the consensus of problems" within subject fields which in turn provided "a sound basis upon which to base the various possible hierarchies of terms" (p. 17). Stated more simply—the "literature organization of concepts *constitutes* the concept of organization of the classification" (p. 17).

Soon after adopting faceted analysis as their preferred methodology the CRG "began to produce a number of individual schemes for such special subjects as soil science, diamond technology, pharmaceuticals, and occupational safety and health" (Foskett, 1971, p.142). In fact, they produced "over 20 special schemes in various areas of knowledge" (Rowley, 1992, p.189). That they dealt mostly with technical and scientific areas "is hardly surprising, given the institutions from which the membership

was drawn—the Department of Scientific & Industrial Research, Tate & Lyle, Metal Box, ICI, the Gas Council and the Patent Office, to name but a few" (McIlwaine and Broughton, 2000, p.196). For instance, B.C. Vickery was initially trained as a chemist, worked as a technical journalist, and then as a librarian at the Ackers Research Laboratory of Imperial Chemical Industries (ICI) for 14 years (East, 1993, p. 855). Douglas Foskett was the librarian at the Metal Box Co., Ltd., as well as the University of London Institute of Education (Harrison, 1993, p. 299).

One important aspect of the CRG's work is that as a group they published little, but as individuals they have been responsible for "a number of milestones in the development of classification theory" (McIlwaine and Broughton, 2000, p.196). They did publish bulletins quite regularly in the 1950s and 1960s that summarized the work they had accomplished and also provided bibliographies of published papers. The 1957 Dorking Conference was another milestone in that it constituted a call for the recognition of faceted classification and for further research to be conducted, as well as bringing the CRG to the "attention of a much wider audience than it had hitherto enjoyed (Foskett, 1964, p.194). It also led to the CRG applying for and receiving a grant of £5000 from NATO to work on formulating a "new scheme of classification of science and technology" that NATO had suggested in their report "Increasing the effectiveness of Western Science" (Foskett, 1969, p.7). This in turn led to another conference, co-sponsored by the Library Association, in 1963. The focus of this meeting was on developing this new general scheme and represented a turning point, in some ways, of the CRG's work.

Much of the historical literature of the CRG points out the influence of Barbara Kyle on this new direction of the CRG's work. Kyle joined the CRG some time in the mid-1950's and according to Furner's citation analysis of the group's work from 1952-2000, her work was most influential in the time period of 1956-62 ( Furner, 2000). Kyle had worked for the Unesco Social Science bibliographies and was a librarian at the Royal Institute for International Affairs. It was her work on social science documentation that led her to produce the Kyle Classification (KC), a faceted classification scheme that dealt with the social sciences as a whole. Kyle's presence had "the effect of extending the interests and discussions of the Group beyond discrete specialized subject fields into the much larger group of disciplines the comprise the Social Sciences" (McIlwaine and Broughton, 2000, p. 196). It was around this time that the members of the CRG begin to see that "the result of the work on special schemes…was really the wrong way to attack the problem" (Foskett, 1971, p.143). This led them to apply for the NATO grant and led to the 1963 conference at which their "research into a new general classification was formally launched" (Austin, 1972, p. 160). The 1958 International Conference on Scientific Information (ICSI) held in Washington with "its immensity of verbiage and its paucity of practical value" also had an effect of refocusing the group's work back toward general classification systems (Foskett, 1962, p.133). Foskett wrote:

> …the CRG turned its thoughts towards a much more complex matter that had received little attention from any of the other schools of thought which had been represented at the two Conferences [Dorking and ICSI]. This is the relation between general and special classifications: is there anything to be gained by pursuing the ideal of a

> new universal classification scheme, and if so, how will the specialist's need be served by it? How can the CRG schemes, for example, that prove so satisfactory for their users, be integrated into such a general scheme? (p. 133)

Theory and practicality in the creation of classification schemes has always been a goal in the work of the CRG and they have done much to fill the gap, so to speak, between the two especially in "such areas as the production of viable formats for use with computers" (Richmond, 1988, p. 246). Once the decision was made to focus on a general classification scheme the group next turned to the theory of integrative levels in the hopes that it would assist them in developing a "bottom-up approach, i.e., forming areas of knowledge after first having pieced together concepts and determining the areas of knowledge they formed" (Spiteri, 1995). Hopkins notes that "it is not possible to determine from published CRG statements whether the group adopted the theory because they accepted it as a true interpretation of the structure of reality, as a true theory of knowledge, or simply as a suggestion (with no judgment as to its truth) for a convenient classification structure" (Hopkins, 1973, p. 206). Regardless, the CRG pursued this avenue but in the process came across several problems with the theory and eventually concluded that it actually "raised more questions than it answered" (Spiteri, 1995). In the end the CRG decided not to use it in the formation of their new scheme. Huckaby's (1972) enquiry into the theory outlines some of the CRG member's interpretation of its meaning and how they tried to adapt it to classification theory. Spiteri (1995) writes that "the significance of the theory is, perhaps, that it provided the CRG with further reinforcement of its belief that areas of knowledge can be determined only after an

analysis of their composition (*a posteriori*), rather than by pre-determining areas of knowledge and then deciding how to break them down into their component parts (*a priori*)." It seems, then, that this experiment with integrative levels, while not producing a new system, did reinforce the CRG's original stance that moving away the old traditional classification's top-down, *a priori* approach was the correct path to take. It did not deter the group, though, from seeking to create a new scheme as it

> occupied the member's attention throughout the 1960s and into the early 70s. The classification per se never saw the light of day, but ideas and discussions of the Group bore fruit in the PRECIS system of indexing devised by Derek Austin and used by the *British National Bibliography* until the advent of computerization brought faster, cheaper, and less labour-intensive (and far less effective) approaches to subject retrieval in a national bibliographic listing (McIlwaine and Broughton, 2000, p.96).

The influence of the CRG on the field of library and information science has been expounded in many of the papers discussed so far in this chapter. McIlwaine and Broughton, each a CRG member during its later years, write that the suggestion at the 1948 Royal Society Scientific Information conference that a committee be formed to study classification issues was to have, in time, "a transforming effect on the theory of knowledge organization throughout the remainder of the century" (McIlwaine and Broughton, 2000, p.195). They reiterate that while the CRG, as a group, did not publish much, the number of individual members' publications "are astronomical" (p. 197). Their influence on the educational aspects of LIS are also quite numerous. At least five

works were published during the 50s, 60s, and 70s that were used as standard works in teaching library students. Many of the members themselves—Vickery, Mills, Foskett, Langridge, Farradane, Morgan, Redfern, Hansen, Cochrane—were teachers (p. 197). Again, McIlwaine and Broughton write that

> these standard works, coupled with the teaching of the fundamental principles that are embodied in them and that are the enduring features of the Group's work, have been transmitted to students for half a century, so that today many teachers as well as students are unaware of the origin of the ground rules which they instill in their pupils (p. 197).

Their influence can also be seen in the continued efforts, lead by original founding members Mills, Coates, and Foskett, of producing the Bliss Bibliographic Classification, edition 2 (BC2). This system "embodies many of the principles developed by the CRG in the creation of special classification schemes and indexing systems during the 1960s and 1970s" (McIlwaine and Broughton, 2000, p.198).

Richmond extols similar virtues of the CRG in her 1988 article entitled "Precedent-Setting Contributions to Modern Classification" which, while focusing on the work of B.C. Vickery, must also acknowledge his connection to the CRG. She writes that the CRG has been notable for many reasons

> …[T]he first is that it has been able to maintain itself over a relatively long time for a small group interested in an intellectual pursuit. It had the good sense to keep a record of its meetings, so that one can trace the gradual development of its views. Secondly, members of the Group produced original, well-organized logical systems, applicable to new or

> revised needs of the various communities which they served. In the third place, its members worked in libraries and information centers where they could innovate and experiment (Richmond, 1988, p. 246).[1]

Hunter notes that the "success of the CRG also led to the formation of other organizations… the best known being the Classification Research Study Group in the United States" (Hunter, 1987). Justice (2001) in examining the history of the "proto-CRG," from 1948 to 1952, describes the group as a "phenomenon of great value" that resulted out of a "certain constellation of bright personalities, modernist but also nonconformists, with a flair for organization and being involved, coalesced in the effervescent zone where scientific work and information work form a unique climate for adventure" (p.1).

Furner (2000) sees the adoption by the CRG of Ranganathan's facet analysis as being "responsible both for developing this theory to the 'standard' state that we take for granted today, and for fully exploring its practical potential by constructing numerous and various real-life schemes, many of which have enjoyed widespread and successful use over more than a quarter century." On the other hand, Furner's citation analysis of the group's work led him to conclude that "the work of the CRG remains undervalued in North America." The results of another citation analysis investigating bibliographic classification literature as a whole reveals that much of the literature of the CRG, from both individual members and collective group work, form the core literature of this particular area of LIS research (Afolabi, 1983, p. 165-167).

In all the historical literature examined, however, there is not much reported about the CRG's views of other methods of information organization and retrieval, particularly

machine-based technology, beyond the words of the members themselves. Foskett (1962), perhaps, said it best when he wrote his report on the activities of the CRG from 1952-1962:

> In all this, I have made no mention of punched cards and all the other hardware. The CRG would have been hard put to it to ignore this, even if it had wanted to, which it does not. We believe, however, that there will, in the foreseeable future, remain a need for classification to provide research workers with the opportunity for browsing and for imposing some discipline on a literature that tends always towards greater disorder. We believe that, since hundreds of millions of dollars and rubles are being spent on hardware, and fat volumes roll off the presses almost day and night, that ten shillings a year that the CRG collects from its members will not be missed (p.137).

The issue of financial resources seemed to have been always a stumbling block to the work of the CRG. Again, Foskett commented that if they had "unlimited financial resources, we should not be bothered by the difficulties that now face us" (p. 133). Coates (1997), in reflecting on classification work in the years after the Dorking conference, wrote that "despite there being successful demonstrations of faceted classification principles in action, interest in them faded in a climate dominated by mechanisation and short-term cost savings on human intellectual work" (p. viii). It is hoped a better understanding of the CRG's views of work outside their own will be demonstrated in the citation analysis.

**Center for Documentation and Communication Research**

The historical literature about the CDCR is not as extensive as that of the CRG and contains little more than superficial reports on its work. Bowles (1999) remarks that "ironically, since its demise in the early 1970's, it [CDCR] has received no historical analysis" (p. 240). Again, similar to the CRG, much of the CDCR's goals and accomplishments have been reported by the members themselves.

In 1955 a very short announcement was placed in the *American Documentation* journal describing the establishment at the School of Library Science at Western Reserve University of the Center for Documentation and Communication Research (CDCR). This announcement described the center as "the first of its kind" and that it would "play a major role in uniting the two professions in a single discipline dedicated to increasing the availability of graphic records" ("A Center for Documentation," 1955, p.178). The two professions in question were library science and documentation (later to be called information science) and the schism that had become apparent between the two fields which was the source of much discussion at the time. Bowles' dissertation focused on this schism in particular and he devoted a large portion of his chapter five to the CDCR and its work and influence on mechanized documentation. One of the major reasons given for the formation of the Center was the recognition of "the growing importance of the field of documentation and the need for personnel adequately trained in its philosophy and techniques" ("A Center for Documentation," 1955, p.178). The 1955 announcement also outlined briefly the goals activities that the Center would undertake:

- Research to define principles and techniques of documentation for the benefit of business and professional men, scientists, scholars, and society in general.

- Contract based service that would meet the needs of specific governmental agencies and industrial enterprises with the objective of improving information retrieval systems and analysis and derivation of practical solutions to individual problems.
- Liaison services that will keep organizations up-to-date on current developments in the field of documentation
- Workshops and conferences to encourage sharing of information and experiences
- Seminars and laboratories to further enrich the education program of the School

Furthermore, they stressed the establishment of the Center would "not only contribute to the solution of problems in bibliographic organization in several academic disciplines represented on the campus, but also that by cooperating with these subject departments the work of the Center can itself be greatly strengthened" ("A Center for Documentation," 1955, p.178).

In October of that same year, *American Documentation* carried a similar bulletin describing in detail a conference held at the Center entitled "The Practical Utilization of Recorded Knowledge--Present and Future" in which the goal of fulfilling the "needs of business, government, research management, scientists, lawyers, doctors, librarians, information specialists, educators, equipment manufacturers, and others concerned with the effective management of recorded information" was again stressed ("Conference on," 1955, p. 255).

Bowles work on the CDCR emphasizes the marketing aspect of the CDCR and does not delve very deeply into the actual nuts and bolts of its work, as does the historical literature on the CRG. He has done extensive background work on uncovering and illuminating the network of techniques the CDCR used to fund its research and market its products. Those products consisted of such things as "specialized information programs

in metallurgy, diabetes, disease vector control, electrical engineering, law, and grammatical language patterns" (Bowles, 1999, p. 242-243). These products were the result of a supremacy in documentation, machine aids, and machine literature searching (p. 240). The CDCR received a "great deal of support from the scientific and technical community" (p. 244) and in return strove to keep these benefactors abreast of their latest developments by sending out newsletters, inviting distinguished guests, sponsoring conferences, and even producing a short film on its mechanical punched card retrieval system designed for disseminating information to metallurgists (p. 263).

Jesse Shera wrote an article on the Center for the *Encyclopedia of Library and Information Science* that emphasized more the actual work of the CDCR, including the work done on semantic factoring and the building of a prototype mechanical searching selector that used punched paper tape (Shera, 1971, 223-224). Shera was the dean at the library school at Western Reserve and it was he, with the help of university president John Millis, that established the Center. He brought in Allen Kent and James Perry to run the CDCR, which they did until Kent and Perry departed in the early 60s. Shera had a vested interest in both the practical and theoretical interests of the Center. In 1961 he wrote a short article entitled "The Librarian and the Machine" for *Library Journal* in which he stressed that the

> …the mechanization of information storage and retrieval has much to contribute to the solution of the library problem, but with it comes the danger of a blind and indiscriminate onrush toward over-simplified solutions and promises of panacea. The overselling of an idea when it is still in its experimental stage will lead to sketchy and ill-defined

> programs, the prostitution of ideals, and a sacrifice of quality to the end that mechanization *per se* may be discredited and condemned for faults that are not inherent in it…. (Shera, 1971, p. S6).

It is not clear, however, if he was referring to work being done at the CDCR. For example, did he believe that the automation being developed there was overtaking the basic principles of information retrieval? It is interesting, though, that he made such a strong statement in the same article that discusses the Center. He cautioned the library search operators to "proceed slowly and ask the right questions at every turn" (p. S6).[2]

Allen Kent himself reported on the work being done at the CDCR and emphasized the division between librarians and documentalists which he described as "not healthy for either" (Kent, 1961, p. 772). Talk such as this is a reoccurring theme in the literature that reports on the work of the CDCR and one can only wonder how much it affected, or guided, their work. This particular article also reported on the "control of terminology and subject headings used to facilitate searching via information retrieval systems that are language based" (p. 773). Graphic records, the documents, were analyzed for aspects that were likely to be of interest to potential users and then these aspects were represented, or expressed, in some way in the language of the retrieval system. Kent cautioned that "an analyst reading a document is tempted to use words found in that document to record the results of his analysis…these words are not necessarily the ones most likely to be used by inquirers" (p. 774). He then described a *multidimensional* approach in which "source documents may be characterized from more than one point of view, and retrieved by combining more than one aspect of subject matter" (p. 775). To do this the techniques of "indexing, classifying, and abstracting"

were used and Kent describes each process in detail. What is interesting is his description of classifying in which he, again, emphasized documentation versus librarianship:

> …in both fields, however, a record can be characterized from more than one point of view in guides to a collection—in a classified catalog for example. The difference here is simply that, in documentation research, more attention is paid to the multidimensional approach than is usual in librarianship. As for "terminology control", this is implicit in the design of classifications. Predictions must be made as to those resemblances or differences which will be useful in literature searches, and a definition of each class made on the basis of the elements that it comprises (Kent, 1961, p.778).

This bears a striking resemblance to the conclusions made by the Classification Research Group drawn from the work they were doing at the same time. However, in no historical literature of the two groups found thus far is this comparison ever made except for one article that B.C. Vickery, of the CRG, wrote for *American Documentation* in 1959 in which he analyzed the semantic coding that Perry, Kent, and Madeline Henderson had developed for the machine literature searching. He concluded that "the semantic code is more in the tradition of Ranganathan's faceted classification and Farradane's [another CRG member] relational operators" (Vickery, 1959, p. 241). He wrote at the end that the "WRU code…deserves to be seriously studied and evaluated. The hard work put in at Cleveland must not be wasted" (p. 241).

After Perry and Kent departed the Center in 1960 and 1963 respectively, the CDCR came under the direction of A.J. Goldwyn and it "continued its research in non-

conventional methods of information retrieval and other aspects of automation, but has, in recent years, broadened its scope to include exploration of library service in a variety of social problems, especially those relating to the inner city and the disadvantaged" (Shera, 1971, p. 226). The Center then moved towards developing new techniques for library routines, as well as investing health and social sciences information retrieval issues. There is not much else written about the further development of the semantic code or its use in machine literature searching. A Case Western Reserve University newsletter, entitled HERE, reported on July 30, 1971 that the CDCR "ceased to be an operating unit within the School of Library Science. The functions which the Center performed over the years have been absorbed by the school and other components of the University" (*HERE*, 1971).

Farkas-Conn (1990) discusses the semantic work of Perry, Kent, and Henderson but oddly enough does not mention the CDCR beyond describing their work at the Western Reserve University and a conference held there that was co-sponsored by the American Documentation Institute (ADI). She does point out that the "Western Reserve group had the political savvy to realize the importance of fostering interest in scientific information among politicians and the industrial community" (p. 187). She, like Bowles, emphasizes more the CDCR's marketability of their products more than the influence of their work on the LIS field. It is expected that a more detailed picture of the CDCR will result from the citation analysis, including their influence and their views of other methods of information organization and retrieval.

One last interesting discovery of documentation of members of both the CRG and the CDCR is the website entitled *Pioneers in Information Science Scrapbook* in which

many members from the LIS community as a whole contributed essays of their most cherished memories of the work they have done, people they had met and ideas they had experienced. Many of the essays from the CRG and CDCR members focus on work done during the late 1950s and early 1960s.

## Summary

This review of the literature about the CRG and the CDCR has shown that the CRG has been more thoroughly researched and documented then the CDCR. The CRG began it work under the contention that the general classification schemes at the time were inadequate for the developing information retrieval mechanisms. Despite a severe lack of funding, they produced a vast amount of research and developed specialized classification schemes that could theoretically be used in the new computerized systems being developed. The CDCR, too, rejected the classification schemes of the times and focused on developing punch card mechanisms and processes that were generously funded by both government and corporate funding. While the CRG tended to look for fundamental principles that would aid classification and retrieval of information, the CDCR were more inclined to develop practical, or pragmatic, methods of retrieval without benefit of good theoretical foundations. To date, there has been no real analysis of the intersection of these two groups' work.

## Endnotes

[1] I did manage to obtain quite a number minutes of the CRG meetings, but was not able to obtain the earliest minutes from meetings held in the 1950s and early 1960s. Hence, I could not trace the development of the views as well as I should have liked.

[2] H. Curtis Wright (1988) wrote an occasional paper on Jesse Shera in which he detailed Shera's eventual disenchantment with the IR research being conducted at the time, as well as the tension that developed between Shera, Perry, and Kent. This tension resulted in Perry and Kent leaving the CDCR in the early 1960s.

# CHAPTER 3

# METHODOLOGY

## Methods and Design

The present investigation employed both qualitative and quantitative research methods to examine more accurately the nature of the relationship between library classification research and information retrieval research, as defined in chapter one, during the stated time period of 1952-1970. An author co-citation analysis (ACA) was used to examine the relationship between the CRG and the CDCR members' published works. The ACA employed a trio of multivariate analyses clustered the authors and then characterized those clusters of authors according to conceptual similarities or differences in their published works. The theory of normative behavior was then applied in order to explain how these similarities and differences were manifested.

It was felt these methods are the most effective way to answer the research questions at the core of this investigation:

1. What was the initial relationship between the Classification Research Group and the Center for Documentation and Communication Research and did this relationship change between 1952 and 1970?
2. What are the conceptual similarities and differences of these two groups as evidenced in their published artifacts from 1952 to 1970?
3. How can we characterize the relationship between the small worlds of the two groups using the theory of normative behavior and its concepts of worldviews, social norms, social types, and information behavior?

To answer the first question it is necessary to identify the geography of the research areas of two groups based on their published works. This can be accomplished

by a citation analysis that will help to construct an overall "map" of the works and then will look at the degree of co-citation among those works. Co-citation analysis is a measure of how often two documents, or authors, are cited together. This in turn will give a hint of how the knowledge base of these two groups was used by the members and others in the field of library and information science at the time. A knowledge base emerges from the process of citing, reviewing and evaluating published works within a discipline by the participants of that discipline and the incorporation of that knowledge into its "educational and professional training programs" (Griffith, 1990, p.42). It is this process of evaluation and incorporation that "seems to be the measure of the importance and stability of the discipline" (p. 42). Understandably, this study focuses on two groups within the larger discipline of LIS and so the "knowledge base" will not be fully investigated, i.e., it will be limited to the library classification research and information retrieval research, but even within those two areas it will have its limits because the CRG and the CDCR produced only a portion of the research conducted in these areas.

The second question is perhaps more difficult to address as it hinges on identifying research fronts and any sort of rapid change of ideas within the two groups--but again, these types of characterizations can be looked for via the author co-citation. The challenge will lie in the interpretation of the results of the analyses. In his longitudinal study on collagen research, Small (1977) stated that "one of the most important and difficult problems of citation studies has been the proper form of validation" (p. 154). He believed that an "independently derived picture" that could be compared with the citation picture was needed to validate his findings. Borgman (2000)

echoes this opinion when discussing the criteria for the inclusion, and exclusion, of papers in a book entitled *Scholarly Communication and Bibliometrics*. She writes

> our editorial discussions led to…accepting only papers that incorporated quantitative analyses of the written record of communication (either bibliographic description or the content of the communication artifact) and a behavioral interpretation of the communication process involved. Thus a large body of bibliometric research was excluded from consideration: that which studied structure without considering the associated processes (p. 145-146).

The third research question, as indicated above, seeks to address this behavioral aspect by applying the theory of normative behavior and its four concepts—social types, social norms, worldviews and information behaviors—to the clusters of authors and the characterization of these clusters that result from the co-citation analysis. Thus, a validation of the "structure" of how the authors' published works were used as revealed by the co-citation analysis can be accomplished by defining the four concepts for each cluster and determining if they coincided with how their scientific communication, i.e., behavior, influenced their work and that of the LIS field in general. For instance, if two clusters share worldviews and information behaviors but are distance from each other (as determined by a statistical analysis) then I could conclude a division in research may have occurred.

Brewer and Hunter (1989) write that the use of a multiple methods of measurement is "simple, but powerful" because its "fundamental strategy is to attack a research problem with an arsenal of methods that have non-overlapping weaknesses in

addition to their complementary strengths" (p. 17). Miles and Huberman (1994), speaking of the argument between qualitative and quantitative proponents, believe that "numbers and work are *both* needed if we are to understand the world" ( p. 40). This study is not proposing to understand the *world*--only a very small portion of it-- but it is with this belief that the following research design is outlined.

The following sections detail the methodology used for data collection and analysis as well as any strengths and weaknesses that warrant mention. Table 3.1 gives an overview of the sequence of the methods.

## Citation Analysis: data collection and processing

As stated above, citation analysis, specifically an author co-citation analysis (ACA), will be used primarily to provide a picture of the subject specialties of the two groups as well as their degree of communication. Citation analysis techniques have been very effective in illuminating the network of papers in specific research areas or fronts. In his 1965 paper entitled "Networks of Scientific Papers" Price likened these networks to knitting in that the rows or strips of stitches "represent objectively defined subjects whose description may vary materially from year to year" and that working out the nature of such strips "might lead to a method for delineating the topography of current scientific literature" (p. 515). Price was referring specifically to citation analysis of journals but this could easily apply to locating a network of published works between the members of specific groups, such as the CRG and the CDCR. Indeed, many studies employing a citation analysis are performed at the level of journals from a discipline as a whole or

**Table 3.1 Overview of data collection and analysis methods[1]**

**Collection of sample of authors from both the CRG and the CDCR**

Identification of authors
- published reports, bulletins, pamphlets, and internal and external correspondence, and conference proceedings

**Analysis of author co-citation data** (McCain, 1990)

Retrieval of citation frequencies
- manually pulled from Science Citation Index (SCI) and Social Science Citation Index (SSCI) printed volumes

Compilation of raw co-citation matrix
- Removal of authors using mean minimum co-citation rate and connection frequencies

Construction of correlation matrix
- Diagonal values indicating self-citation removed and replaced with adjusted value

Multivariate Analysis of Correlation Matrix
- Factor analysis, cluster analysis, and multidimensional scaling (MDS) using SPSS 11.0
- Creation of author cluster map, cluster characterization, and cluster intercorrelation

Interpretation and Validation
- Inspection
- Application of concepts taken from theory of normative behavior
  - Define social types, social norms, information behavior , worldviews for each cluster.
  - Compare with proximity of clusters on author map and their intercorrelations
  - Draw conclusions about research correlations of CRG and CDCR based on this comparison

identify the major influential persons in that field or specialty (Price 1965; see also Small & Griffith, 1974; Griffith, Small, et al, 1974a; White & Griffith, 1982 ; McCain, 1983, 1984, 1986a, 1986b, 1989, 1990; Bayer, et al., 1990; Paisley, 1990; Braam, et al., 1991a, 1991b; Fazel & Danesh, 1995; van Raan & van Leeuwen, 2001). But, it has also been suggested that "the journal is too broad a unit of analysis to reveal the fine structure of specialties" (Small and Griffith, 1974, p.18). For this reason, an author co-citation analysis (ACA) looking at individual authors and their works is the primary goal.

Lievrouw (1990) argues that citation analysis reveals the communication between researchers in a field because the strict convention of referencing others' work can be

"construed to stand for interaction among the authors and those they cite" (p. 61). The technique of "mapping" the network of citations is an effective method used to show this communication. The mapping can illustrate graphically how authors and their works have been cited over time, i.e., the strengths of those citations, who is citing them and how often. This in turn helps to illuminate the core papers of a research area. Core papers are defined as those cited most often over a given period of time. To some degree mapping may show how the papers are being used. This is based on the type of paper in which another is cited. For example, a theoretical paper cited in an experimental paper may be an indication of the testing and acceptance of that theory. White (1990) writes that ACA "reveals useful order hidden in the author data of a bibliographic database. It is a way of algorithmically classifying authors' *oeuvres* so that they strongly imply subjects" (p. 430).

Similarly, Griffith (1990) argues that "there are very strong social organizations underlying scientific work" and, with the co-citation research conducted by himself and in collaboration with Henry Small (Small & Griffith, 1974;Griffith, Small, et. al, 1974) he states that what is of major importance is the "discovery of a bibliographic information structure that parallels social and intellectual structure" (p. 44). He further states that "the discovery of rapid changes in that [bibliographic information] structure instantly integrated Kuhn's revolutionary groupings and changes in Price's modeling of the overall community" (p. 44). Indeed, Small (1977) found a "physical and measurable manifestation of a collective mental switch from a static to a dynamic conception of collagen" in his landmark longitudinal study, but he also debated on whether or not it was a "revolution" in Kuhn's sense. He did conclude that "a true Kuhnian paradigm, if one

exists, would manifest itself through citation data in precisely the way exemplified by collagen" (p.159). With this in mind, however, this portion of the study is limited to the tasks as described by White and McCain (1998) so eloquently in the following statement: "All ACA can do, for the historian of ideas or any other party, is to identify influential authors and display their interrelationships from the citation record" (p. 327).

**Sample of authors**

The members of the CRG and the CDCR have been identified through an investigation of published reports, bulletins, pamphlets, internal and external correspondence, minutes, and conference proceedings (Table 3.2). The CRG published regular bulletins (Classification Research Group, 1956a, 1956b, 1958, 1959, 1961, 1962, 1964, 1968) in the *Journal of Documentation* during the 1950s and 1960s in which they often provided a list of member names as well as bibliographies of any recent member publications. These bulletins were the primary source of names for the sample and were cross-referenced with reports (CRG 1969) and conference proceedings ("Proceedings," 1957; "Proceedings," 1959). CDCR members were more difficult to identify as they did not publish regular bulletins in any of the more widely recognized journals. James Perry and Allen Kent, along with Jesse Shera, edited several series of volumes that were either proceedings from the various conferences held at the Center or were simply collections of papers first published in the journal *American Documentation*; see Shera, Kent, & Perry, (1956, 1957); Perry & Kent (1957, 1958); Kent (1960-1961). Contributors to these publications were cross-referenced with CDCR brochures and internal reports found in the CDCR archived papers at the Case Western Reserve University Archives, as well as

periodically published bibliographies. From these sources a final list of members was composed.

Because the time span of the study covers eighteen years it was decided to include those authors who were members of the two groups at any time within this period, regardless of the actual length of their membership (e.g., Vickery drifted away from the CRG in the early 1960s and Perry and Kent moved on from the CDCR at approximately the same time. However, I have included their work across the entire time span of the study). The reasoning behind this choice is the belief that work begun while a member of the group can certainly be continued beyond the confines of the group, as well as work started or published before becoming associated with either group. Furthermore, the citing of an author's work in the intervening years may have resulted directly from their association with either group. From a practical standpoint, it must also be noted that a significant sample of authors would most likely not be possible if the citation data was restricted to only those years the authors were confirmed members. The life spans of published artifacts is also an important factor. For instance, Price (1965) found "an indication that about half the bibliographic references in papers represent tight links with rather recent papers, the other half representing a uniform and less tight linkage to all that has been published before" (p. 514). One of the goals of this study is to try to identify any research fronts coming from the two groups, hence the time span and the decision to look at all published works cited during this time. Again, Price lends support by suggesting that "…in the special circumstances of being able to isolate a 'tight' subject field, we find that half the references are to a research front of recent papers and that the other half are to papers scattered uniformly through the literature" (p. 515).

The other justification for the sample extent comes directly from the second research question of this study which seeks to identify conceptual similarities and differences of these two groups, as evidenced in their published artifacts. As such, this study is deliberately aimed at the *oeuvres* of the identified authors. White and Griffith (1981) define this to mean "the body of writings by a person--and not the person himself" (p. 163). Obviously, due to the eighteen year limitation, the *oeuvres* will not be comprehensive as many of the authors published before 1952 and well after 1970. But, this research seeks to understand both the writings, and to some extent the authors, in an attempt identify the behavioral aspects which Borgman (2000) concluded were needed in any truly comprehensive bibliometric study. Indeed, White and Griffith (1981) state a similar stance when explaining the limitations of their own study:

> *Oeuvres*, we conclude, are inherently ambiguous objects of study. To uncover why they appear on the map as they do, one must make a detailed examination of the writings in them and of how and why they are cocited with writings by other authors. This, of course, involves an analysis at the level of specific documents, which is beyond our present scope, but may be possible in other contexts. (p. 166)

The present scope of this study is also limited, but will attempt an analysis of the overall general content of the published works where possible. A detailed analysis of the works at the specific documents level will be discussed in the Future Research chapter.

**Retrieval of cocitation frequencies.** The procedure for the ACA comes directly from McCain's (1990) paper detailing the steps for performing the analysis on a set of known authors. This procedure has been used effectively by McCain in a series of

previous studies; see McCain (1984; 1986a; 1986b; 1988, 1991) and White and McCain (1998). Dr. McCain was also directly consulted concerning this procedure, both in person and via email communications.

Publications for each author were retrieved using the *Social Science Citation Index* (SSCI) and the *Science Citation Index* (SCI) for the period of the mid 1950s to the mid 1960s. Because of the time period in question, print-based citation manuals were used and the information was manually collected.[2] These actual citation indexes were ten year cumulative manuals beginning in 1945 and extending up to 1970.

Each author identified through the methods discussed previously was located in the manuals and citation information was retrieved. A typical citation entry supplies the author's name (last name, first initial) and the abbreviated name of the journal in which the work appears, year of publication, volume and pagination. Underneath this entry is the name of the person citing that work, the journal in which that work appears, publication date, volume and pagination. A typical entry for author Jason Farradane:

**Farradane J**
**61 J Doc 17 233**
**Costello JC J Chem Doc 3 164 63**

Farradane's 1961 article appearing in the *Journal of Documentation*, volume 17, beginning on page 233 has been cited in 1963 by Costello in the *Journal of Chemical Documentation*, volume 3, beginning on page 164.

Every time a work is cited a name will appear, i.e., some entries often had long lists of citers which can be interpreted as a clear sign of the influence of the work. The title of the actual cited work is never supplied in the citation index unless it is a book or conference and even then it is abbreviated. However, SSCI and SCI also provide a

Source Index which does list the titles of works. As such, both the citation index and the source index were consulted and cross-referenced when necessary.[3] The indexes also provide a list of journals used in the index so if an abbreviated journal title was not clear it was easily deciphered by consulting this list. This was often a necessary step when collection of the cited works commenced as will be discussed later.

One of the inconsistencies encountered in the culling the citation indexes were the various forms of names encountered. For instance, Jason Farradane, a CRG member, is indexed in several different ways: Farradane; Farradane, J.; Farradane, JE; Farradane, JEL. Furner (2001) encountered this same problem while querying the online version of the indexes. When naming inconsistencies were encountered all forms of the name were verified and any citation data included with them were used.

At this stage self-citations—an author citing her or himself—were collected and are dealt with at a later stage in the methodology. Furthermore, because all papers are listed by the first author only in the indexes the question of determining who is the first named in co-authored papers is safely answered (Garfield, 1979, p. 242). However, it does present a limitation. For instance, if a member of one group co-authored a paper with a member of the other group I would not know that relationship existed because of the incompleteness of the index.

**Compilation of raw data matrix**

Each author's citation information was entered by hand into a separate Excel spreadsheet with only the citing information entered under the author's name. McCain indicated that scanning the actual citation index entry had been tried on earlier occasions but proved too difficult to convert electronically to a spreadsheet (personal

conversation, 2001). Including only the citing information is deliberate as it allows for the combining of all the spreadsheets into one master spreadsheet which can then be downloaded into a SPSS statistical software program. To do this a value of "1" is entered in the cell accompanying each citing entry. For example, figure 3.1 is a portion of a spreadsheet for CRG member Brian Vickery:

**Table 3.2 Typical spreadsheet showing citing information for Brian Vickery**

| Citing doc, ISI print format | Vickery, B.C. |
|---|---|
| anthony lj rep pr phys 32 709 69 | 1 |
| artandi s special lib 57 571 66 | 1 |
| ashmore wsh j libr 1 253 69 | 1 |
| atherton p lib res tec 9 463 65 | 1 |
| averbukh vm nau t inf 2 1970 29 70 | 1 |
| baker nr am documen 19 363 68 | 1 |
| bartlett lc j am med a 199 244 67 | 1 |

Each entry must be in the prescribed form as illustrated above in Table 3.1. This allows for the running of a macro function, supplied by Dr. McCain, called "Flatten Data" that combines all the citation information from each separate spreadsheet into a master. The macro is run and "flattens", or compresses, the data. The procedure itself is time consuming as it requires patiently copying and pasting one spreadsheet at a time into the master spreadsheet, running the macro and then repeating the procedure. In the end, the spreadsheet consists of author's names in columns and citers in rows alphabetically. Table 3.3 shows a small slice of the master spreadsheet:

**Table 3.3 Example of master spreadsheet showing combined citation data**

| **Citing doc, ISI print format** | AITCHSON | ARNOLD | BAGLEY | BOOTH | BROXIS |
|---|---|---|---|---|---|
| snodey sr ieee e writ ews 22 64 | 1 | 1 | 0 | 1 | 0 |
| soergel d inf storage 3 219 67 | 0 | 0 | 0 | 1 | 0 |
| ???? j doc 20 166 64 | 0 | 0 | 0 | 0 | 0 |
| ???? j doc 12 227 56 | 0 | 0 | 0 | 0 | 0 |
| ???? rev int doc 32 72 65 | 0 | 0 | 0 | 0 | 0 |
| fairthor ra ann r info 4 73 69 | 0 | 0 | 0 | 0 | 0 |
| bourne cp ann r info 1 171 66 | 1 | 0 | 0 | 0 | 0 |
| richmond pa coll res li 27 23 66 | 1 | 0 | 0 | 0 | 0 |
| chernyi ai nau t inf 2 6 68 | 1 | 0 | 0 | 0 | 0 |
| king dw ann r info 3 61 68 | 0 | 0 | 0 | 0 | 0 |
| sastri mi met inf med 7 49 68 | 0 | 0 | 0 | 0 | 0 |

A value of 1 indicates a citation and a value of zero indicates no citation. As illustrated here Aitchison, Arnold and Booth are all cited by S. R. Snodey in a 1964 publication. The "Flatten Data" macro essentially matches the citers and combines the 1 values into one row thereby showing cocitation occurrences. Citers lacking names are indicated by "????"--these can be anything from published bibliographies or anonymous editorials. Once all the citation data has been entered into the spreadsheet and compressed the final outcome shows all the co-citation occurrences between authors. Total cites are tallied for each row and column, made into hard values so the entire spreadsheet can then be sorted in descending order by the total cites row. Any papers in the **Citing doc, ISI print format** column that cite only one author are removed from the spreadsheet. Similarly, any author not found to be cited with any other author is removed as well. Finally, the individual paper column is deleted.[4] This final spreadsheet is the matrix that will be downloaded into the statistical software program SPSS.

This entire cocitation frequency retrieval process is essentially the same as the online procedures illustrated in various bibliometric studies (Sandstrom, 1998; White &

McCain, 1998; McCain, 1984; 1986a; 1986b; 1988, 1991; White & Griffith, 1981; Bayer, Smart & McLaughlin, 1990; Culnan, O'Reilly, & Chatman, 1990 ). In these studies the researchers discuss the process of querying the online SSCI and SCI databases using various query statements and building their database electronically. The only difference with the process outlined here is that it is completed manually from start to finish and requires considerable more typing and eye-strain. The final product, however, is the same—a matrix of author cocitation frequencies that allows for the next step in the ACA process.

**Construction of correlation matrix.** McCain (1990) states that the "first step in mapping or clustering cocited authors is the conversion of the raw data matrix to a matrix of proximity values, which indicate the relative similarity or dissimilarity of author-pairs" (p. 435). The resulting correlation matrix then simply displays all the "possible combinations of correlation values for a set of variables" (Hernon, 1994, p. 154). In this case the variables are the authors themselves. McCain (1990) reports that some researchers will calculate the Pearson correlations at this stage of the analysis or they will calculate them at a later stage of the multivariate analyses. The Pearson correlation, or Pearson r, measures the strength of the linear relationships between two variables and "reflects the extent to which variations in one variable accompany variations in the other variable" (Hernon, 1994, p. 153). In the case of Pearson r, the measurement is always between -1 and +1. A score falling in this range will indicate a positive or negative relationship, e.g., "-1 means a perfect negative, +1 a perfect positive relationship and 0 means the perfect absence of a relationship" (Pearson Product Moment Correlation, n.d.). Kreuzman (1990) found in his study of the relationship between contemporary

epistemology and philosophy of science that "a high positive Pearson r for a pair of philosophers is interpreted as intellectual closeness, while a high negative Pearson r is interpreted as intellectual distance" (p.61). However, he does caution that care is needed when using the terms 'intellectual closeness' and 'intellectual distance' because these two things are not a measure of the similarity of views, but rather of "their work and subject matter" (p. 61). Using a similar approach for this study, the Pearson r will be calculated and used as an indication of the strength of the relationship between author co-citation pairs. These two correlations raise two very important questions. First, if two authors have a negative relationship—an indication of intellectual distance—but share similar work styles and subject matters, can this be used as an indication of a gap in the relationship of the two subfields as discussed in the introduction? Second, do high positive relationships correspond to the work styles and subject matters of those authors cited often together? These two questions will be addressed later in the study.

The raw data matrix in the Excel spreadsheet was opened up into a data editor in the SPSS 11.0 software program with the names of the authors in columns as variables. A frequency analysis was run to check for missing values and then the data was run again with a multiple response analysis. This analysis allowed for the production of cross-tabulations -- authors' names are entered in identically ordered rows and columns. Once the cross-tabulation was completed the matrix was re-entered by hand into a new Excel spreadsheet. The diagonal values in this spreadsheet are an indication of self-citation and will skew the co-citation results if not dealt with by blanking them out and replacing them with an adjusted value. Table 3.4 shows an example of a cross-tabulated matrix with diagonals value bolded. From this example it is observed that Cleverdon has cited

himself seventy-seven times, Aitchison twenty-two times, Casey nineteen times and Booth six times.

**Table 3.4 An example of cross-tabulation and diagonal values**

| | AITCHSON | BOOTH | CASEY | CLEVRDON |
|---|---|---|---|---|
| **AITCHSON** | **22** | 1 | 2 | 16 |
| **BOOTH** | 1 | **6** | 3 | 1 |
| **CASEY** | 2 | 3 | **19** | 2 |
| **CLEVRDON** | 16 | 1 | 2 | **77** |

The next step is to calculate the average citation rate for each author and replace the blanked out diagonals with this value. This process is dealt with differently in various studies. White and Griffith (1981) decided to scale the values by "taking the three highest intersections and dividing by two …which would approximate the next highest score in the distribution" (p. 165). McCain (1990) elected to treat "the diagonal cells values as missing data and calculating cocited author correlations accordingly" (p. 435). Sandstrom (1998) used the unique character string 999 for her co-citation analysis of anthropologists. Paisley ( 1990) simply blanked the values out altogether. McCain (1990) notes that whatever value is used in the diagonal there is little to suggest a "difference in mapping, clustering, and factor analysis between scaling the diagonal values and treating them as missing data" (p. 435).

For this study the diagonals values will be replaced with an average co-citation for each author. So, using the example in figure above the resulting matrix would appear as in Table 3.5.

**Table 3.5 Diagonal values are replaced with average co-citations**

| | AITCHSON | BOOTH | CASEY | CLEVRDON |
|---|---|---|---|---|
| **AITCHSON** | 2.83 | 1 | 2 | 16 |
| **BOOTH** | 1 | 0.96 | 3 | 1 |
| **CASEY** | 2 | 3 | 1.83 | 2 |
| **CLEVRDON** | 16 | 1 | 2 | 6.83 |

McCain (1990) observes two advantages to creating a correlation matrix: (1) for any given pair of authors the matrix functions as a "measure…of how similar their 'cocitation profiles' are" and (2) it "removes differences in 'scale' between authors who are highly cited and those….who are less frequently cited overall" (p. 436). Once the correlation matrix is completed it should be run through another frequency analysis so as to "debug it," that is, to insure no variables are missing (McCain, personal conversation, 2001). At this point it is then ready to be used in a series of multivariate analyses.

**Citation Analysis: Multivariate Analysis of Correlation Matrix**

A combination of three types of analyses—factor analysis, cluster analysis and multidimensional analysis—were used to display the inter-author relationships within the similarity matrix. The SPSS 11.0 software program was used to perform these analyses. All three methods are very useful and provide illustrations, both statistically and graphically, of how the authors and their works communicated and interrelated during the particular twenty year time period in question. It is very much like creating a multi-level picture or collage and then dissecting it layer by layer. One could even liken it to an archaeological excavation—the dig site is marked off and then is very patiently excavated down to an agreed upon level or to the bedrock itself. Every item found within the parameters is carefully cataloged and its position noted with relation to every other item

in the pit. These relationships can go up or down, from side to side, at diagonals, etc., and some may not even relate at all, which is an interesting phenomenon in and of itself. With each successive excavated layer a more complete picture of the pit can be drawn. The trick of it lies in the interpretation and inferences gleamed from these relationships between these items. In a study such as this ACA the somewhat open interpretations are generally accepted analytical risks as it has been shown that similar end results and conclusions can result from differing analytical approaches. McCain (1990) notes as well that "a number of studies have been more or less intuitive accounts of the researchers' own subject areas, and the interpretations based on personal knowledge" (p. 441).

**Cluster Analysis**

The first method is a cluster analysis that is basically empirically forming "clusters or groups of highly similar entities" (Aldenderfer & Blashfield, 1984, p. 7). In a range of citation studies this is the more widely used method of identifying clusters of core authors or core journals (Small, 1976,1980, 1981, 1986; Small & Crane, 1979 ; Small & Griffith, 1974; Griffith & Small, et. al., 1974; McCain, 1984; 1986a; 1986b; 1988, 1991; White & McCain, 1998 ).

There are five basic steps in a typical cluster analysis, according to Aldenderfer and Blashfield (1984):

1) selection of a sample to be clustered
2) definition of a set of variables on which to measure the entities in the sample
3) computation of the similarities among the entities
4) use of a cluster analysis method to create groups of similar entities
5) validation of the resulting cluster solution

The first three steps will have been completed previous to the cluster analysis, and the creation and validation of the cluster can be accomplished by using SPSS software. McCain (1990) has found that the most used approaches to cluster formation are the

> 'hierarchical agglomerative' vs. 'iterative partitioning'—essentially a 'bottom-up building' versus a 'top-down splitting' of clusters. In the former, individuals and/or groups of individuals are gradually joined, and then those clusters joined in still larger clusters. In the latter the single cluster of all individuals is split, as are subsequent clusters, until only individuals remain. ACA research has tended to use the agglomerative clustering approach (p. 437).

Hierarchical agglomerative clustering is one of seven families of clustering techniques and is one of the most frequently used of these seven. Aldenderfer and Blashfield (1984) speculate that this technique is more popular because it is conceptually simple to understand. It uses a simply N x N similarity matrix (here N refers to the number of authors) which merges all the most similar cases, and it requires "exactly N-1 steps to cluster a similarity matrix" (p. 36). Lastly, it produces a very nice visual representation called a dendrogram that displays the hierarchical organization. See Figure 3.1 for a typical dendrogram.

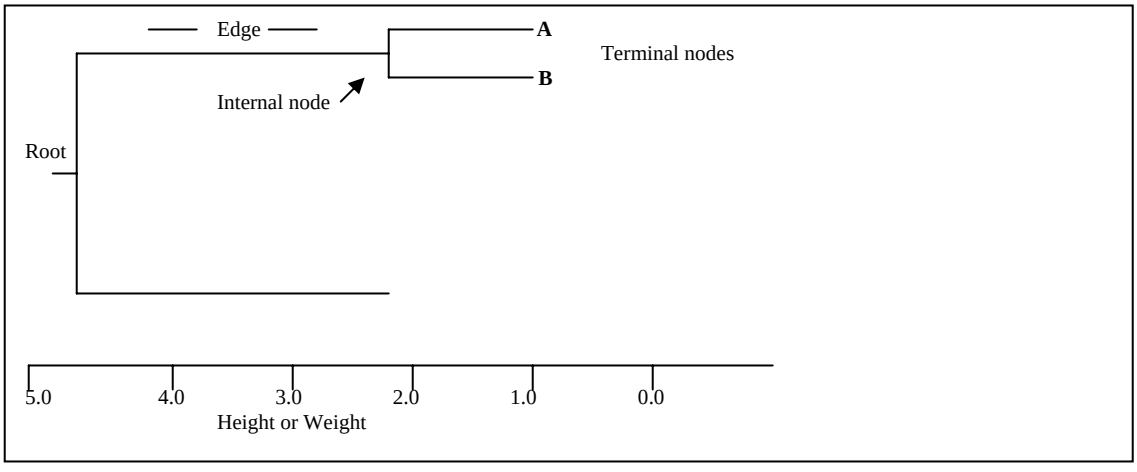


**Figure 3.1 A typical dendrogram adapted from Everitt, Landau, & Leese, 2001.**

A dendrogram is simply a mathematical and pictorial representation of the complete clustering procedure. The nodes represent the clusters and "the length of the stems represent the distances at which clusters are joined" (Everitt, Landau, & Leese, 2001, p. 70).

To obtain these results there are several *linkage* procedures within the technique that can be used. For an ACA, complete linkage and Ward's method are the "best" linkage procedures in that they both "give similar and interpretable results on the same data set" (McCain, 1990, p.437). For this study a complete linkage agglomerative clustering technique will be employed. This is also sometimes called "furthest neighbor" in that it tends to find compact clusters with maximum distance between objects (Everitt, Landau, & Leese, 2001, p. 63). Basically, this means that "any candidate for inclusion into an existing cluster must be within a certain level of similarity to all members of that cluster" (Aldenderfer and Blashfield, 1984, p. 40). It is hoped this kind of linkage should complement well any conclusions made about the overall characteristic of the cluster.

The interpretation of the dendrogram will involve determining the number of clusters of authors. McCain (1990) notes there are "no generally accepted stopping rules to guide the researcher in selecting the best set of clusters to report" (p. 437). Aldenderfer and Blashfield (1984) say the obvious question is where to cut the nested tree structure of the dendrogram so that the "optimal number of groups is found" and that sometimes it can be accomplished by “subjective inspection of the different levels of the tree” (p.53-54). Everitt, Landau, and Leese (2001) write that "to use the results involves choosing the partition, and the best way of doing this is unclear" (p. 89).

**Multidimensional Scaling**

The second method that will be used is multidimensional scaling. This technique involves the correlation matrix of proximities used for the cluster analysis in order to generate a *spatial* representation, i.e., a map of the configuration of points usually shown in two dimensions (Kruskal & Wish, 1978, p. 7). This in turn reveals the structure of the similarity or dissimilarity of the objects being analyzed. Objects that are dissimilar will be further apart on the map and objects that are similar will be closer together on the map (p. 7).

The maps generated are especially useful for a citation analysis because "the points on the map represent the individual authors placed according to their interauthor similarities" (McCain, 1990, p. 438). Typically, maps show the author groups (akin to "schools") in the field study and their location with respect to each other, the centrality and peripherality of authors with groups and with respect to the overall field, proximities of authors within groups and across group boundaries, and position of authors with respect to the map's axes (White & Griffith, 1981, p. 165). The dimensionality of the maps "refers to the number of coordinate axes…a direction of particular interest in a configuration, or some underlying characteristic of the objects under study" (Kruskal & Wish, 1978, p. 48). For example, an ACA map can show how clusters of authors are oriented along a horizontal or vertical axis that represents such things as their subject interests or style of work which are two axes that McCain (1984) found in her study of macroeconomics literature. Sandstrom's (1998) study of anthropologists found within the clusters of authors dimensions ranging from sexual selection to foraging theory along the vertical and horizontal axes. In his study of theories of rationality in epistemology

and philosophy of science Kreuzman (1990) found in his ACA a vertical axis "roughly indicating the willingness to employ…quantitative or mathematical models in understanding rationality" (p. 72). The horizontal axis represented a subject axis "ranging from philosophy of science on the left to epistemology on the right" (p. 72). Most ACA studies involve a two-dimensional interpretation because it captures "a high proportion of the variance (85% or more) in the proximities matrix and provide rich matter to interpret" (McCain, 1990, p. 439).

In the conversion of the correlation matrix to a two-dimensional map, however, there is some loss, or distortion, of information. For instance, White and McCain (1997) found the "two-dimensional space in which the authors appear is relative, not absolute, and it fails to capture certain relationships among oeuvres that appear in higher dimensionality" (p. 331). There are a variety of methods to measure the distortion such as a "stress" test which can determine the "best fit between the original input matrix 'distances' and the estimated distances in the chosen low-dimensional solution" (McCain, 1990, p. 438). Stalans (1995) explains that a stress level is measured from 0 to 1, with small values (e.g., 0 to .15) indicating a good fit, 0 indicating a best fit, and larger values indicating a bad fit (p. 149). The R Square ($R^2$ or RSQ) is another measure of the proportion of variance and is also measured on a scale from 0 to 1, but with higher values (those closer to 1) indicating a better fit. McCain (1990) writes that "since author co-cited data are inherently "noisy," a higher stress value (but usually less than 0.2) is considered an acceptable trade-off for a two- or three-dimensional solution if the R Square is high" (p. 438). Both of these measures will be used in the analysis of the data of this study.

The ALSCAL program in SPSS will be used to generate a two-dimensional map of the interauthor relations using the correlation matrix previously mentioned. This program will generate both a stress measurement and an R Squared (RSQ) measurement, as well as a coordinates for plotting the authors in a Euclidean space (i.e., an two-dimensional plane with x, y coordinates for each author). A scatterplot will also be produced. This diagram "provides a visual picture of the relationship between proximity data and MDS-computed distance data" (Stalans, 1995, p. 155). Distances are shown on the horizontal axis and proximities on the vertical. If the relationship between the two is *smooth*, i.e., "data points are close together and large gaps are absent" then there is good relationship between distance and proximity. Figure 3.2 shows a typical scatterplot diagram.

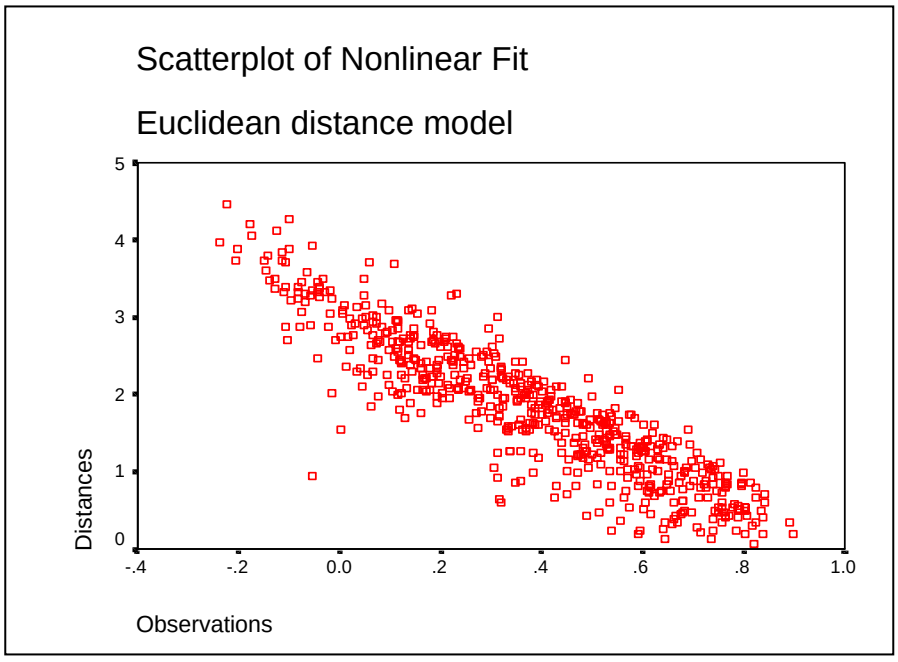



**Figure 3.2 A typical scatterplot produced by ALSCAL**

The interpretation of the visual representation, as stated earlier, was "based on author placement and author cluster orientations along the horizontal and vertical axes" (McCain, 1990, p. 439). The actual author map resulting from this study and it analysis will be discussed in chapter four.

**Factor Analysis**

The last method of multivariate analysis performed was a factor analysis which explored the relationships between authors (variables) in the correlation matrix in order to "represent a set of variables in terms of a small number of hypothetical variables" (Kim & Mueller, 1978, p. 9). The positive relationships in the matrix can be illuminated by discovering subsets of variables and then further analyzing the relationships between those subsets with the goal of "ascertaining the minimum number of hypothetical factors that can account for the observed covariation" (p. 9). For this study an exploratory factor analysis was used. This means that I explored the co-citation data in order to reduce it and make some interpretations of the relationships between the authors' works. Kline (1994) suggests rephrasing this approach to ask "what constructs or dimensions could account for the correlations" between the variables (Kline, 1994, p.7). Kline (quoting Royce, 1963) states further that a "factor is a constant operationally defined by its factor loadings" (p. 5). Factor loadings are the correlation of the variable with the factors. Here, each author (variable) is correlated with the subset of variables that result. White and Griffith (1982) found that "the size of the author's loadings on a factor measures identification with that speciality in the eyes of the citers. The number of factors on which an author loads—that is, his or her factorial complexity—is a measure of cross-speciality usefulness" (p.259). McCain (1990) writes that determining how authors may

contribute to more than one factor amounts to determining the "author's breadth" (p. 440).

The goal of the analysis is to produce a *simple structure* which is "a condition in which variables load at near 1 (in absolute value) or at near 0 on an eigenvector (factor). Variables that load near 1 are clearly important in the interpretation of the factor and variables that load near 0 are clearly unimportant" (Bryant & Yarnold, 1997, p. 132). In order to produce this simple structure the factors must be rotated, since the rotation method being the most important decision in any factor analysis. Kline (1994) writes that much of the scientific value of factor analysis depends on proper rotation" (p. 64). Rotation simply involves rotating the factors on axes; by doing so one may theoretically obtain "an infinity of mathematically equivalent sets of factors" (p. 56). Just as determining the number of clusters needed in the cluster analysis requires a stopping point, so also a stopping point is required for determining the number of factors needed. In this case a *scree test* is used, which is the "best solution to selecting the correct number of factors" (p. 75). A graph is made of the eigenvalues, a mathematical property of the matrix. The cutoff point of factors occurs where the slope of the line changes. "The factors with eigenvalues that lie on the path of steep descent … are retained [and] the factors with eigenvalues that come later…are not extracted" (Bryant & Yarnold, 1997, p. 132).

For this study, a oblimin rotation method was calculated to generate a structure matrix of the main components or factors. Oblimin is a type of oblique rotation that "tries to simplify the pattern matrix by way of reference axes" (Kim & Mueller, 1978, p. 78).

Again, the intercorrelations found among the authors is the goal of this analysis and McCain (1990) writes that

> the strength of the intercorrelation, if any, among the factors may reveal subject-related linkage *above the author level*. If factors are uncorrelated ("independent"), the subject specializations they represent may not yet have been linked by citers—an important insight. An oblique factor rotation (oblimin as opposed to varimax) will suggest whether they are independent by providing a matrix of factor intercorrelations. In highly coherent fields, certain factors may have intercorrelations of 0.3 and above, pointing to links between research specialties or other constructs (p. 440).

Citation analysis studies may use either orthogonal or oblique rotation methods. McCain (personal correspondence, 2002) explains that the method chosen depends on what you want to look at and how you think the world might be structured. For instance, White & Griffith (1982) chose an oblique rotation method for their factor analysis of studies in science, technology, and society (SSTS). They used a sample of seventy-one authors and obtained co-citation data for each as well as "counts of the co-occurrence of every cited author's name with selected, broadly connotative terms from the titles of the citing papers" (p. 260). Using sixteen title terms and the author correlations found loaded on seven factors, and "to throw the main structure into relief, we show only the loadings above .40 for authors and above. 60 for title terms" (p. 262). What is most important to note is the judgmental labels they placed on each of the seven factors—philosophy, social history, specialty structure, social psychology, communication, policy, and economics.

This labeling is performed after the authors with the highest loadings for each factor are "boxed in rank order" because "authors with boxed loadings in common make immediate sense as subject specialties" and together with the title terms White and Griffith felt their judgmental labels were validated. Setting the limit for loadings is an important step in the analysis. For instance, McCain (1984) used only loadings above +/-0.5. Authors may also load on several factors and "the magnitude of author's various loadings is of interest in estimating his or her different impacts" (White & Griffith, 1982, p. 264). Lastly, the intercorrelations of the factors themselves may be taken into account.

In essence, all this rather complicated computation—graciously performed by the SPSS software—allows the researcher to name the factors, i.e., identify the specialty of a group of authors and how the authors "fit" into that specialty. This is very complementary to the two previously discussed types of analysis—cluster analysis and multidimensional scaling. The final product of these three analyses will be an author map showing the clusters of authors which will then be "characterized." This characterization will involve, to a certain degree, an examination of the cited works of the authors involved. This will be discussed further in the interpretations and validations portion of the methodology, as well as in the chapter on future research.

**Interpretations and Validation**

The three types of multivariate analysis used will provide what McCain (1990) calls "complementary, frequently reinforcing results" (p. 442). The interpretation of the three multivariate analyses will rely in great part on "discovering what the author clusters, factors, and map dimensions represent in terms of scholarly contributions, institutional or geographical ties, intellectual associations and the like" (p. 441). For

example, White and McCain (1997) performed an exhaustive domain analysis of information science covering the years 1972 to 1995. They examined eleven different aspects of the field—disciplinary and institutional affiliations, specialty structure of the discipline, canonical authors, axes on which authors mapped, etc.—using the same methods described in this chapter. They were able to draw several important conclusions about the make-up of the field of information science based on just the statistical evidence gathered and their interpretation of the results. They also to some degree examined some of the actual literature of the authors in question, specifically to be able to put a name on the varying clusters of authors and to determine if a paradigm shift had occurred during the 1980s. Using a similar approach, I will also undertake a small analysis of the literature published by the two groups being studied here in order to verify the findings of my statistical analysis. This will also assist in the naming of the components that result from the factor analysis and in the identification of axes via the multidimensional scaling.

However, for this study it is felt that an additional form of interpretation and validation is needed to provide what Borgman (2000) described as the “behavioral interpretation of the communication process involved” (p. 146). This communication process, as discussed at the beginning of this chapter, is the strict convention of referencing others' work which then can be "construed to stand for interaction among the authors and those they cite" (Lievrouw, 1990, p.61). Thus, this study seeks to find out if citation behavior, as an indication of this communication process, can be deciphered in much the same way as social behavior is in any number of interactive information exchanges. I have employed a parallel case study of two groups of these two groups of

researchers during an eighteen year time period with the simple intent of discovering how their work either came together or diverged as it pertains to classification research and information retrieval research. To do this I have sought to use these concepts from Chatman's theory of normative behavior as a means of answering this question.

This theory has grown out of many years of Chatman's research into the small worlds of different groups of people and their use of information in many types of social situations. For instance, Chatman has investigated information use among the working poor, elderly women, women in prison, janitors, and before her death in 2002 she was studying a group of people known as "dirt eaters," i.e., they ate dirt. Chatman's many investigations into information and public behaviors has involved her using several different theories, such as *diffusion theory*, *alienation theory*, *gratification theory*, and *social network theory* (Chatman, 1986, 1990, 1991, 1992, 1996, 1999). From these studies she developed three of her own theories: *theory of information poverty*, *theory of life in the round*, and the *theory of normative behavior* (Chatman 1996, 1999; Burnett, Besant, & Chatman, 2001). It is the last theory that I seek to use in this study, in large part because of a study Chatman conducted with Burnett and Besant on virtual communities and feminist booksellers that led them to conclude that the theory "provides…a reasonable conceptual strategy for examining and evaluating both the place of information within a social world and the socially valued interactions between people and individuals and information in that world" (Burnett, Besant, & Chatman, 2001, p. 545). This observation led me to believe the theory could be applied in some ways to this study.

Specifically, application of this theory to this study involved thinking in terms of the *small worlds* of traditional classification researchers and information retrieval researchers. A s*mall world*, as defined in chapter one, is “a specific context that serves a particular population to permit its members to conduct their business in a routine, expected manner" (Burnett, Besant, & Chatman, 2001, p. 536). Very simply, the *theory of normative behavior* seeks to explain “the ways in which people deal with information in the contexts of their small worlds" (p. 536). The four concepts used with the theory—social types, social norms, worldview, and information behavior—have already been previously defined and discussed in chapter one, so I will only lightly readdress them here. Social types "pertains to the classification of a person or persons, and are the absolute definitions given to individuals within a social world" (p. 537). Social norms help to hold a small world together and "give individuals a way to gauge what is 'normal' in a specific context at a specific time" (p. 537). Worldview is a "collective perception held in common by members of a social world regarding those things that are deemed important or trivial" (p. 537). Finally, information behavior is "a state in which one may or may not act on available or offered information" (p. 537).

This study also relies heavily on the idea of scholarly communication—the social processes by which scholars “use and disseminate information through formal and informal channels” (Borgman, 2000, p. 144). In particular, it focuses on the formal channels of communication between scholars via the written record, but also consider, to some degree, the background of these scholars as it has shaped and influenced this work. By background is meant such things as their education, employment, personal and institutional influences, and interactions with others in the field, etc. Using this, the

question can then turn to applying the four concepts discussed above to the quantitative evidence resulting from the author co-citation analysis.

In their study of virtual communities and feminist booksellers, Burnett, Besant, and Chatman took each of these four concepts and applied them to their observations of the two groups. They begin with clearly defining the small world of each group, which I have also done with classification researchers and information retrieval researchers. Second, they took each concept in turn and discussed in detail the evidence for each within the groups. For example, there are two main "social types" in virtual communities: "insiders and outsiders, those who are members of the community and those who are not" (Burnett, Besant, & Chatman, 2001, p. 543"). Because this is a virtual community where communication is based on writing, the participants—insiders—are mainly "valued according to their writing abilities" and are "often judged based on other characteristics of their writings, including the level of personality coming through the text, frequency of postings, the appearance of self-revelation…" and the like (p. 544). Social norms among feminist booksellers were found to center around "space which is created specifically by and for women," as well as emphasizing women over men in terms of employment and wages (p. 543). Lastly, cooperation, rather than competition, was more highly valued. Feminist booksellers were also found to gravitate toward the worldview that "women have been oppressed by the culture of patriarchy" (p. 543). Information behavior in virtual communities consist mainly of "making announcements, posting information queries and replies, providing links to outside information sources, etc." (p. 545). These are just a few examples how the *theory of normative behavior* has been applied to two specific groups of people. The researchers found the "ways in which these concepts

manifest themselves can vary widely according to the specific make-up of a given social world" (p. 545).

## Summary

This chapter has covered the methodology of this study. The methodology consists of a citation analysis plus a method of validation. The citation analysis was described in terms of data collection and processing techniques; and it also covered three types of multivariate analysis: cluster analysis, multidimensional scaling, factor analysis. The author cluster map and characterization of those clusters by factor analysis and multidimensional scaling will be taken as evidence of the "make-up" of the small worlds of the two groups, just as White and McCain did in their 1997 study of information science. Lastly, this evidence will then be further interpreted by the application of the four concepts of the theory of normative behavior as discussed in the interpretation and validation section above. A validation of the findings should result. The research questions will then be re-examined and answered in the next chapter and following that final conclusions will be presented.

## Endnotes

[1] This table is modeled after a similar table in Sandstrom's (1998) dissertation (p. 145). Her model, in turn, is based on McCain's (1990) technical overview of ACA.

[2] I could not electronically collect the data needed because the online citation index available to me did not go back to the years in question

[3] I collected source index information as well in order to perform a bibliometric coupling analysis as future research. It should be noted, however, that not all entries in the citation index were mirrored in the source index, which begs the question of the overall accuracy of the indexes.

[4] Obviously successive changes to the spreadsheet demands much copying and pasting of the data into new spreadsheets so as to avoid "erasing" what was done previously, i.e., if a mistake is discovered it is easy enough to retrace the steps in the procedure and correct it. As such, each time a new step is performed a new spreadsheet is created. Needless to say this is a good practice to follow in any scientific analysis, much akin to keeping a notebook in a chemistry experiment.

# CHAPTER 4

# DATA RESULTS AND ANALYSIS

## Sample and Selection of authors

As described in the methodology the sample of authors used in the author co-citation analysis were selected through an examination of published papers, reports, bulletins, pamphlets, internal and external correspondence, minutes, and conference proceedings from both the Classification Research Group (CRG) and the Center for Documentation and Communication Research (CDCR). From this forty-three (43) authors were identified from both groups combined. Table 4.1 gives a complete list of authors, listed by group.

Using the Science Citation Index (SCI) and the Social Science Citation Index (SSCI) citation data between the years 1952 to 1972 was located for each other author. Citation data (cites) consisted of a citing author, the journal citing, and year of publication. However, it was not possible to identify data for every author as they either did not publish as first authors during those years or they published in journals that were not covered by the two indices. From the CRG, no citation data was found for Campbell, Finerty, Jones, McIlwaine, Pendleton, and Watkins. This left only twenty (20) authors. For the CDCR, no citation data was found for Alvin, Cort, and Hazelton. This left only fourteen (14) authors. Table 4.2 lists these remaining thirty-four (34) authors.

**Table 4.1 Original sample of authors identified. N=43**

| Classification Research Group (CRG) | Center for Documentation and Communication Research (CDCR) |
|---|---|
| Aitchison, J. | Alvin, J. |
| Arnold, D.V. | Booth, A.D. |
| Bagley, D | Casey, R.S. |
| Broxis, P.F. | Cort, D.E. |
| Campbell, D.J. | Egan, M.E. |
| Cleverdon, C. | Goffman, W. |
| Coates, E.J. | Goldwyn, A.J. |
| Fairthorne, R.A. | Hazelton, R. |
| Farradane, J.E.L. | Kent, A. |
| Finerty, E. | Melton, J.L. |
| Foskett, D.J. | Melton, J.S. |
| Jolley, J.L. | Overmeyer, L. |
| Jones, G. | Perry, J.W. |
| Kyle, B.R.F. | Rees, A. |
| Langridge, D. | Saracevic, T. |
| McIlwaine, I.C. | Zull, C. |
| Mayne, A.J. | |
| Mills, J. | |
| Morgan, T.S. | |
| Palmer, B.I. | |
| Pendleton, O.W. | |
| Rippon, J.S. | |
| Watkins, K.H. | |
| Vickery, B.C. | |
| Wells, A.J. | |
| Whitrow, M. | |

Citation data for the remaining thirty-four (34) authors was collected into a master spreadsheet containing 952 cites. This was then further refined by removing any papers citing only one author, leaving 181 cites with co-citations ranging from eleven (11) authors to two (2) authors. This final master spreadsheet, considered the raw data matrix,

was used in the construction of a correlation matrix to be used in the multivariate analysis.[1]

**Table 4.2. Author sample after citation data collected. N=34**

| CRG | CDCR |
|---|---|
| Aitchison, J. | Booth, A.D. |
| Arnold, D.V. | Casey, R.S. |
| Bagley, D. | Egan, M. |
| Broxis, P.F. | Goffman, W. |
| Cleverdon, C. | Goldwyn, A.J. |
| Coates, E.J. | Kent, A. |
| Fairthorne, R.A. | Melton, J.L. |
| Farradane, J.E.L. | Melton, J.S. |
| Foskett, D.J. | Overmeyer, L. |
| Jolley, J.L. | Perry, J.W. |
| Kyle, B.R.F. | Rees, A. |
| Langridge, D. | Rees, T.H. |
| Mayne, A.J. | Saracevic, T. |
| Mills, J. | Zull, C. |
| Morgan, T.S. | |
| Palmer, B.I. | |
| Rippon, J.S. | |
| Vickery, B.C | |
| Wells, A.J. | |
| Whitrow, M. | |

## Author Exclusion and Final Correlation Matrix

The decision of whether or not to eliminate authors who are deemed not statistically significant enough to include in the multivariate analysis, i.e., those whose co-citation mean falls below a certain cut-off, is treated differently across many ACA studies. McCain (1990) writes that some ACA researchers may select "only those

authors meeting certain ad hoc criteria, such as mean co-citation rates above nine (for 10 years of Social Scisearch data) and co-citation with at least one-third of the entire author set" (p. 435). Sandstrom (1998) performed an ACA covering nine years and 45 authors and in the end kept 35 authors with high co-citation as well as some low-count and marginally connected authors because she was interested in "understanding how both core and peripheral information is discovered and used by contributors to a specialty literature" (p. 433). However, Sandstrom also found that including low-count authors who have "virtually no correlation with the other authors…produced highly erratic communality values" (p. 433). She found that some experimentation with including and then excluding some authors was necessary and so I have done the same.

Keeping in mind the eighteen year time span of this study, the criteria for this exclusion was based on an examination of average co-citations resulting from the frequencies and multiple response analyses, as well as the original raw data matrix of citation counts. In particular, if an author was determined to have a low co-citation count, but was cited among the most highly citing papers then I felt justified in keeping that author. This also involved "counting 1s and 0s" in the cross tabulations matrix, discussed below; 1 meaning an author was only cited once with any other author, 0 meaning they were never co-cited with anyone (McCain, personal conversation, 2001). McCain suggested removing any author who has mainly 1s and 0s in their column in the cross tabulations matrix. The frequency of the citations comes to bear as well. For example, Melton (J.L.) and Booth were two authors I felt, after experimenting with the different analyses, needed to remain in the matrix, despite having low-counts. Melton had a total of twenty-two (22) and Booth a total of twenty-six (26) co-citations, with an

average co-citation of .65 and .76 respectively. In the end, thirteen (13) authors were excluded from the analysis: Arnold (10), Bagley (3), Broxis (7), Egan (1), Langridge (22), Mayne (8), Morgan (8), Overmeyer (6), Rees, T. (3), Rippon (3), Wells (17), Whitrow (10), and Zull (7). The numbers in parenthesis are their co-citation counts. Egan was the lowest and Langridge the highest. I did consider keeping both Langridge and Wells, but found that among the four low-count or marginal authors—Booth, Langridge, Melton, and Wells—keeping these two would produce erratic results in the multivariate analyses. Table 4.2 shows the final sample of twenty-one authors, deliberately mixed, with range of years in which citation data was found for each.

**Table 4.3 Final sample of authors used for the multivariate analyses. N=21**

| | |
|---|---|
| Aitchison, J. (1961-1963) | Kent, A. (1955-1967) |
| Booth, A.D. (52-69) | Kyle, B (52-67) |
| Casey, R. (58-63) | Melton, J.L. (58-62) |
| Cleverdon, C. (54-70) | Melton, J.S. (58-68) |
| Coates, E.(53-69) | Mills, J. (52-67) |
| Fairthorne, R. (52-70) | Palmer, B. (53-68) |
| Farradane, J. (52-69) | Perry, J. (52-69) |
| Foskett, D. (52-70) | Rees, A. (57-68) |
| Goffman, W. (63-69) | Saracevic, T. (58, 64-69) |
| Goldwyn, A. (62-67) | Vickery, B. (52-70) |
| Jolley, J.(55-68) | |

The construction of the final correlation matrix involved entering the raw data matrix, stripped of both the citers' column, the total cites column, and with non-

significant authors removed, into the SPSS 11.0 program. Both a frequencies analysis and a multiple response analysis were run. The frequencies analysis generated a summary of co-citation frequencies for each author, as well as ensuring there were no missing values.[2] The multiple response analysis generated a cross tabulations matrix which was then used as the basis for the correlation matrix. The cross-tabulations were re-entered by hand into a fresh Excel spreadsheet with authors names in both row and column headings. As discussed in the methodology, the diagonals (showing author self-citation counts) are typically blanked and replaced with individual author co-citation averages calculated from the remaining values. Tables 4.3 and 4.4 show the correlation matrix with the original diagonals and the correlation matrix with re-calculated diagonals respectively.

## Results: Multivariate Analyses of Correlation Matrix

The final correlation matrix found in Table 4.4 was then subjected to the three multivariate analyses as discussed in the methodology chapter. The results of each analysis is first described separately and then as a whole.

### Cluster Analysis

The hierarchical cluster analysis yielded very interpretable results. Figure 4.1 below shows the dendrogram as produced from the hierarchical cluster program in SPSS 11.0. Initially, the twenty-one authors were clustered into two large clusters, with five authors—Goffman, Saracevic, Goldwyn, Melton(J.S.), and Rees—comprising the smaller cluster and the remaining eighteen authors comprising the larger cluster. I chose to

**Table 4.4 Correlation matrix with original diagonals**

| | AITCHSON | BOOTH | CASEY | CLEVRDON | COATES | FAIRTHRN | FARADANI | FOSKETT | GOFFMAN | GOLDWYN | JOLLEY | KENT | KYLE | MELTONJL | MELTONJS | MILLS | PALMER | PERRY | REESA | SARCEVIC | VICKERY |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AITCHSON | 22 | 1 | 2 | 16 | 1 | 6 | 3 | 2 | 2 | 2 | 2 | 7 | 4 | 1 | 2 | 1 | 0 | 3 | 2 | 0 | 5 |
| BOOTH | 1 | 8 | 3 | 1 | 1 | 2 | 1 | 2 | 2 | 0 | 1 | 2 | 1 | 2 | 1 | 0 | 0 | 2 | 0 | 0 | 3 |
| CASEY | 2 | 3 | 19 | 2 | 1 | 3 | 1 | 1 | 0 | 0 | 2 | 6 | 1 | 2 | 2 | 0 | 1 | 11 | 0 | 0 | 7 |
| CLEVRDON | 16 | 1 | 2 | 68 | 1 | 21 | 7 | 2 | 14 | 4 | 4 | 12 | 3 | 1 | 7 | 3 | 0 | 5 | 18 | 5 | 19 |
| COATES | 1 | 1 | 1 | 1 | 14 | 6 | 5 | 5 | 0 | 0 | 2 | 4 | 2 | 1 | 2 | 5 | 3 | 3 | 0 | 1 | 11 |
| FAIRTHRN | 6 | 2 | 3 | 21 | 6 | 44 | 7 | 6 | 9 | 3 | 3 | 11 | 2 | 2 | 5 | 3 | 1 | 5 | 7 | 3 | 20 |
| FARADANI | 3 | 1 | 1 | 7 | 5 | 7 | 26 | 4 | 3 | 0 | 2 | 7 | 3 | 1 | 1 | 2 | 2 | 5 | 2 | 1 | 10 |
| FOSKETT | 2 | 2 | 1 | 2 | 5 | 6 | 4 | 23 | 2 | 0 | 3 | 6 | 4 | 1 | 1 | 5 | 4 | 2 | 0 | 0 | 13 |
| GOFFMAN | 2 | 2 | 0 | 14 | 0 | 9 | 3 | 2 | 28 | 8 | 0 | 3 | 3 | 0 | 5 | 0 | 0 | 1 | 17 | 6 | 4 |
| GOLDWYN | 2 | 0 | 0 | 4 | 0 | 3 | 0 | 0 | 8 | 15 | 0 | 4 | 1 | 0 | 5 | 1 | 0 | 1 | 8 | 2 | 6 |
| JOLLEY | 2 | 1 | 2 | 4 | 2 | 3 | 2 | 3 | 0 | 0 | 6 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 4 |
| KENT | 7 | 2 | 6 | 12 | 4 | 11 | 7 | 6 | 3 | 4 | 1 | 41 | 4 | 2 | 5 | 1 | 0 | 9 | 7 | 2 | 14 |
| KYLE | 4 | 1 | 1 | 3 | 2 | 2 | 3 | 4 | 3 | 1 | 1 | 4 | 16 | 1 | 3 | 1 | 1 | 2 | 5 | 0 | 5 |
| MELTONJI | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 1 | 2 | 1 | 3 | 1 | 0 | 0 | 2 | 0 | 0 | 3 |
| MELTONJS | 2 | 1 | 2 | 7 | 2 | 5 | 1 | 1 | 5 | 5 | 0 | 5 | 3 | 1 | 16 | 2 | 1 | 4 | 6 | 2 | 8 |
| MILLS | 1 | 0 | 0 | 3 | 5 | 3 | 2 | 5 | 0 | 1 | 1 | 1 | 1 | 0 | 2 | 15 | 3 | 0 | 0 | 0 | 9 |
| PALMER | 0 | 0 | 1 | 0 | 3 | 1 | 2 | 4 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 3 | 9 | 1 | 0 | 0 | 8 |
| PERRY | 3 | 2 | 11 | 5 | 3 | 5 | 5 | 2 | 1 | 1 | 1 | 9 | 2 | 2 | 4 | 0 | 1 | 29 | 2 | 0 | 15 |
| REESA | 2 | 0 | 0 | 18 | 0 | 7 | 2 | 0 | 17 | 8 | 1 | 7 | 5 | 0 | 6 | 0 | 0 | 2 | 39 | 11 | 8 |
| SARCEVIC | 0 | 0 | 0 | 5 | 1 | 3 | 1 | 0 | 6 | 2 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 11 | 16 | 1 |
| VICKERY | 5 | 3 | 7 | 19 | 11 | 20 | 10 | 13 | 4 | 6 | 4 | 14 | 5 | 3 | 8 | 9 | 8 | 15 | 8 | 1 | 70 |

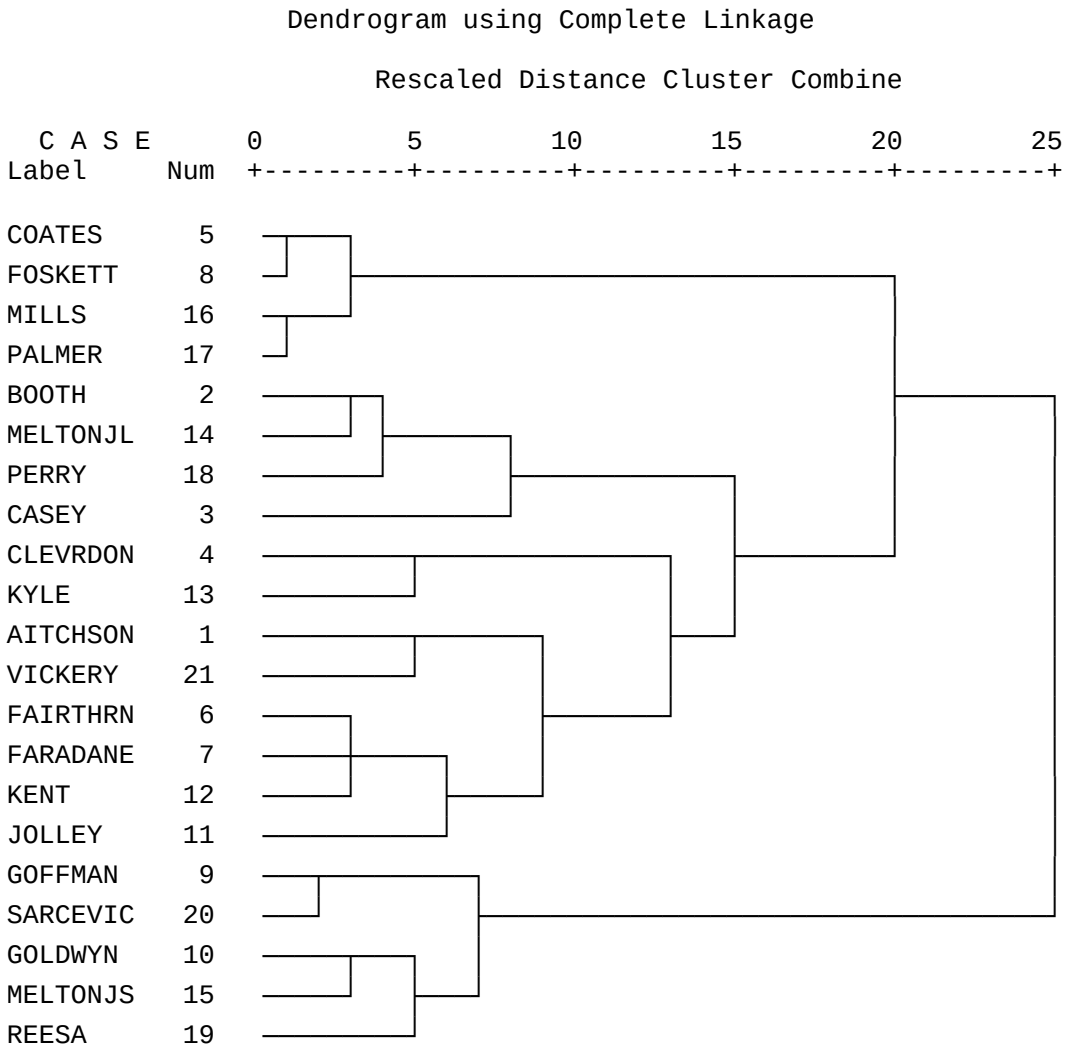


**Table 4.5 Final correlation matrix with diagonals replaced with average co-citations**

| | AITCHSON | BOOTH | CASEY | CLEVRDON | COATES | FAIRTHRN | FARADANI | FOSKETT | GOFFMAN | GOLDWYN | JOLLEY | KENT | KYLE | MELTONJL | MELTONJS | MILLS | PALMER | PERRY | REESA | SARCEVIC | VICKERY |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AITCHSON | 3.1 | 1 | 2 | 16 | 1 | 6 | 3 | 2 | 2 | 2 | 2 | 7 | 4 | 1 | 2 | 1 | 0 | 3 | 2 | 0 | 5 |
| BOOTH | 1 | 1.25 | 3 | 1 | 1 | 2 | 1 | 2 | 2 | 0 | 1 | 2 | 1 | 2 | 1 | 0 | 0 | 2 | 0 | 0 | 3 |
| CASEY | 2 | 3 | 2.25 | 2 | 1 | 3 | 1 | 1 | 0 | 0 | 2 | 6 | 1 | 2 | 2 | 0 | 1 | 11 | 0 | 0 | 7 |
| CLEVRDON | 16 | 1 | 2 | 7.25 | 1 | 21 | 7 | 2 | 14 | 4 | 4 | 12 | 3 | 1 | 7 | 3 | 0 | 5 | 18 | 5 | 19 |
| COATES | 1 | 1 | 1 | 1 | 2.7 | 6 | 5 | 5 | 0 | 0 | 2 | 4 | 2 | 1 | 2 | 5 | 3 | 3 | 0 | 1 | 11 |
| FAIRTHRN | 6 | 2 | 3 | 21 | 6 | 6.25 | 7 | 6 | 9 | 3 | 3 | 11 | 2 | 2 | 5 | 3 | 1 | 5 | 7 | 3 | 20 |
| FARADANI | 3 | 1 | 1 | 7 | 5 | 7 | 3.35 | 4 | 3 | 0 | 2 | 7 | 3 | 1 | 1 | 2 | 2 | 5 | 2 | 1 | 10 |
| FOSKETT | 2 | 2 | 1 | 2 | 5 | 6 | 4 | 3.15 | 2 | 0 | 3 | 6 | 4 | 1 | 1 | 5 | 4 | 2 | 0 | 0 | 13 |
| GOFFMAN | 2 | 2 | 0 | 14 | 0 | 9 | 3 | 2 | 3.95 | 8 | 0 | 3 | 3 | 0 | 5 | 0 | 0 | 1 | 17 | 6 | 4 |
| GOLDWYN | 2 | 0 | 0 | 4 | 0 | 3 | 0 | 0 | 8 | 2.25 | 0 | 4 | 1 | 0 | 5 | 1 | 0 | 1 | 8 | 2 | 6 |
| JOLLEY | 2 | 1 | 2 | 4 | 2 | 3 | 2 | 3 | 0 | 0 | 1.5 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 4 |
| KENT | 7 | 2 | 6 | 12 | 4 | 11 | 7 | 6 | 3 | 4 | 1 | 5.35 | 4 | 2 | 5 | 1 | 0 | 9 | 7 | 2 | 14 |
| KYLE | 4 | 1 | 1 | 3 | 2 | 2 | 3 | 4 | 3 | 1 | 1 | 4 | 2.35 | 1 | 3 | 1 | 1 | 2 | 5 | 0 | 5 |
| MELTONJI | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 0 | 0 | 1 | 2 | 1 | 1.05 | 1 | 0 | 0 | 2 | 0 | 0 | 3 |
| MELTONJS | 2 | 1 | 2 | 7 | 2 | 5 | 1 | 1 | 5 | 5 | 0 | 5 | 3 | 1 | 3.15 | 2 | 1 | 4 | 6 | 2 | 8 |
| MILLS | 1 | 0 | 0 | 3 | 5 | 3 | 2 | 5 | 0 | 1 | 1 | 1 | 1 | 0 | 2 | 1.85 | 3 | 0 | 0 | 0 | 9 |
| PALMER | 0 | 0 | 1 | 0 | 3 | 1 | 2 | 4 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 3 | 1.3 | 1 | 0 | 0 | 8 |
| PERRY | 3 | 2 | 11 | 5 | 3 | 5 | 5 | 2 | 1 | 1 | 1 | 9 | 2 | 2 | 4 | 0 | 1 | 3.7 | 2 | 0 | 15 |
| REESA | 2 | 0 | 0 | 18 | 0 | 7 | 2 | 0 | 17 | 8 | 1 | 7 | 5 | 0 | 6 | 0 | 0 | 2 | 4.7 | 11 | 8 |
| SARCEVIC | 0 | 0 | 0 | 5 | 1 | 3 | 1 | 0 | 6 | 2 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 11 | 1.7 | 1 |
| VICKERY | 5 | 3 | 7 | 19 | 11 | 20 | 10 | 13 | 4 | 6 | 4 | 14 | 5 | 3 | 8 | 9 | 8 | 15 | 8 | 1 | 8.65 |

further break down the larger second cluster into three sub-clusters. This decision was based on three observations. First, the goal of ACA is "generally not to determine the "true" number of clusters in a the matrix, but to inform a more general discussion" (McCain, 1990. p.437). Second, I made a subjective inspection of the tree's different levels, one of the most basic methods, and made a decision as to where to cut off the clusters (Aldenderfer & Blashfield, 1984, p. 54). Thirdly, I cross checked my subjective observation with the 'scree test' from the factor analysis, which I will discuss later. Dr. McCain indicated this was a handy method for determining the number of clusters in a dendrogram (personal conversation, 2001).

The first cluster from this larger second cluster is comprised of Coates, Foskett, Mills, and Palmer. The second is composed of Booth, Melton (J.L.), Perry, and Casey. The third is the largest and is composed of Cleverdon, Kyle, Aitchison, Vickery, Fairthorne, Farradane, Kent, and Jolley. However, it is observed that even these could be broken down into still smaller clusters, but for the purposes of this study, four clusters will suffice.




Cluster 1: Coates, Foskett, Mills, Palmer

Cluster 2: Booth, Melton (J.L.), Perry, Casey

Cluster 3: Cleverdon, Kyle, Aitchison, Vickery, Fairthorne, Farradane, Kent, & Jolley

Cluster 4: Goffman, Saracevic, Goldwyn, Melton (J.S.), Rees (A.)

**Figure 4.1 Dendrogram produced from SPSS 11.0 hierarchical cluster analysis using complete linkage.**

The first cluster is composed of only CRG members. The second and fourth clusters are composed only of CDCR members. The third is composed of seven CRG members and one CDCR member, Kent. The almost complete division of the two groups is interesting, but this in not entirely surprising as it seems to support the initial observation, discussed in chapter one, that these two areas of research were regarded as separate during the time period in question. However, a deeper analysis is needed before any final conclusions can be drawn.

For instance, comparing the dendrogram clusters to the proximity matrix composed of Pearson correlation values illustrates how highly correlated each member is with the other members of their respective clusters and sheds some light on the reliability of the dendrogram. This will be useful when later comparing the cluster analysis with the factor analysis and multidimensional scaling. Table 4.5 shows this proximity matrix.

Cluster one members are highly correlated with each other—scores ranging in the .7-.9 range—but each member is also correlated with two members of cluster three: Farradane and Jolley. Coates and Foskett are rather highly correlated to Farradane, Mills to Jolley, and Palmer, at slightly lesser values, can be correlated to either Farradane or Jolley. Cluster two members are also fairly highly correlated with each other with values ranging from .5 to .8. Appearing again, cluster three member Farradane is correlated with Booth, Casey, and Melton (J.L.). Perry, though, is not surprisingly more correlated with Kent, but again Farradane follows closely behind. However both Kent and Farradane show higher correlation with Perry than Casey who is the least correlated of the entire cluster with values hovering in the .5 range. Cluster four correlations range from .5 to .8 and, like clusters one and two, are connected to cluster three. Saracevic,

## Table 4.6 Proximity matrix of Pearson correlation values

| | AITCHENSON | BOOTH | CASEY | CLEVRDON | COATES | FAIRTHRN | FARADANE | FOSKETT | GOFFMAN | GOLDWYN | JOLLEY | KENT | KYLE | MELTONJL | MELTONJS | MILLS | PALMER | PERRY | REESA | SARCEVIC | VICKERY |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AITCHENSON | 0 | .214 | .251 | .337 | 0.132 | 0.784 | 0.643 | 0.214 | 0.514 | 0.296 | 0.606 | 0.673 | 0.37 | 0.314 | 0.615 | 0.226 | -0.042 | 0.388 | 0.596 | 0.278 | 0.667 |
| BOOTH | .214 | 0 | .575 | .234 | 0.405 | 0.355 | 0.501 | 0.418 | -0.276 | 0.59 | 0.457 | 0.489 | 0.29 | 0.8 | 0.206 | 0.268 | 0.355 | 0.726 | 0.006 | 0.207 | 0.243 |
| CASEY | .251 | .575 | 0 | .198 | 0.43 | 0.32 | 0.565 | 0.404 | -0.184 | 0.019 | 0.254 | 0.506 | 0.22 | 0.766 | 0.311 | 0.167 | 0.25 | 0.553 | -0.068 | 0.245 | 0.394 |
| CLEVRDON | .337 | .234 | .198 | 0 | 0.315 | 0.538 | 0.543 | 0.36 | 0.537 | 0.754 | 0.292 | 0.639 | 0.67 | 0.239 | 0.68 | 0.205 | 0.118 | 0.393 | 0.45 | 0.565 | 0.289 |
| COATES | .132 | .405 | .430 | .315 | 0 | 0.419 | 0.708 | 0.911 | -0.187 | 0.003 | 0.59 | 0.488 | 0.35 | 0.555 | 0.23 | 0.797 | 0.882 | 0.58 | -0.131 | -0.286 | 0.41 |
| FAIRTHRN | .784 | .355 | .320 | .538 | 0.419 | 0 | 0.801 | 0.504 | 0.452 | 0.561 | 0.661 | 0.772 | 0.64 | 0.375 | 0.763 | 0.562 | 0.387 | 0.606 | 0.623 | 0.369 | 0.501 |
| FARADANE | .643 | .501 | .565 | .543 | 0.708 | 0.801 | 0 | 0.794 | 0.168 | 0.302 | 0.734 | 0.769 | 0.57 | 0.609 | 0.614 | 0.678 | 0.535 | 0.637 | 0.318 | 0.088 | 0.69 |
| FOSKETT | .214 | .418 | .404 | .360 | 0.911 | 0.504 | 0.784 | 0 | -0.205 | 0.098 | 0.576 | 0.434 | 0.38 | 0.566 | 0.326 | 0.794 | 0.803 | 0.616 | -0.025 | -0.228 | 0.327 |
| GOFFMAN | .514 | -.276 | -.184 | .537 | -0.187 | 0.452 | 0.168 | -0.205 | 0 | 0.661 | 0.148 | 0.475 | 0.39 | -0.2 | 0.68 | -0.031 | -0.244 | -0.001 | 0.615 | 0.85 | 0.317 |
| GOLDWYN | .296 | .059 | .019 | .754 | 0.003 | 0.561 | 0.302 | 0.098 | 0.661 | 0 | -0.045 | 0.399 | 0.62 | -0.079 | 0.794 | 0.073 | -0.004 | 0.221 | 0.701 | 0.81 | 0.094 |
| JOLLEY | .606 | .457 | .254 | 0.292 | 0.59 | 0.661 | 0.734 | 0.576 | 0.148 | -0.045 | 0 | 0.73 | 0.42 | 0.551 | 0.292 | 0.698 | 0.557 | 0.567 | 0.039 | -0.054 | 0.595 |
| KENT | .673 | .489 | 0.506 | 0.639 | 0.488 | 0.772 | 0.769 | 0.434 | 0.475 | 0.399 | 0.73 | 0 | 0.63 | 0.595 | 0.719 | 0.494 | 0.378 | 0.682 | 0.364 | 0.263 | 0.654 |
| KYLE | .374 | .288 | 0.221 | 0.667 | 0.35 | 0.664 | 0.574 | 0.38 | 0.389 | 0.617 | 0.422 | 0.625 | 0 | 0.261 | 0.551 | 0.422 | 0.347 | 0.449 | 0.222 | 0.455 | 0.339 |
| MELTONJL | .314 | .800 | 0.766 | 0.239 | 0.555 | 0.375 | 0.609 | 0.566 | -0.2 | -0.079 | 0.551 | 0.595 | 0.26 | 0 | 0.225 | 0.363 | 0.4 | 0.794 | -0.145 | -0.329 | 0.351 |
| MELTONJS | .615 | .206 | 0.311 | 0.68 | 0.23 | 0.763 | 0.614 | 0.326 | 0.68 | 0.794 | 0.292 | 0.719 | 0.55 | 0.225 | 0 | 0.302 | 0.167 | 0.481 | 0.706 | 0.608 | 0.437 |
| MILLS | .226 | .268 | 0.167 | 0.205 | 0.797 | 0.562 | 0.678 | 0.794 | -0.031 | 0.073 | 0.698 | 0.494 | 0.42 | 0.363 | 0.302 | 0 | 0.882 | 0.486 | 0.014 | -0.146 | 0.382 |
| PALMER | -.042 | .355 | 0.25 | 0.118 | 0.862 | 0.387 | 0.535 | 0.803 | -0.244 | -0.004 | 0.557 | 0.378 | 0.35 | 0.4 | 0.167 | 0.882 | 0 | 0.508 | -0.188 | -0.217 | 0.173 |
| PERRY | .388 | .726 | 0.553 | 0.393 | 0.58 | 0.606 | 0.637 | 0.616 | -0.001 | 0.221 | 0.567 | 0.682 | 0.45 | 0.794 | 0.481 | 0.486 | 0.508 | 0 | 0.086 | -0.064 | 0.325 |
| REESA | .596 | .006 | -0.068 | 0.45 | -0.131 | 0.623 | 0.318 | -0.025 | 0.615 | 0.701 | 0.039 | 0.364 | 0.22 | -0.145 | 0.706 | 0.014 | -0.188 | 0.086 | 0 | 0.566 | 0.172 |
| SARCEVIC | .278 | -.207 | -0.245 | 0.565 | -0.286 | 0.369 | 0.088 | -0.228 | 0.85 | 0.81 | -0.054 | 0.263 | 0.46 | -0.329 | 0.608 | -0.146 | -0.271 | -0.064 | 0.566 | 0 | 0.162 |
| VICKERY | .667 | .243 | 0.394 | 0.289 | 0.41 | 0.501 | 0.69 | 0.327 | 0.317 | 0.094 | 0.595 | 0.654 | 0.34 | 0.351 | 0.437 | 0.382 | 0.173 | 0.325 | 0.172 | 0.162 | 0 |

Goldwyn, and Melton (J.S.) share connections to Cleverdon. Rees and Goffman are correlated with Aitchison and, in Goffman's case, more so than his link with Saracevic. Melton (J.S.) is actually more highly correlated with Cleverdon, Fairthorne, and Kent (all from cluster three) than with Goffman, Saracevic, and Rees.

These correlations with cluster three is not surprising as it is the central cluster and was the most difficult to judge when it came to deciding the cut off point for the number of clusters as a whole. Cluster three also has the widest range of correlation values, ranging from .2 to .8 and all its members has correlations with outside members falling within that range.

**Multidimensional scaling**

Figure 4.2 below shows the author map of twenty-one authors as calculated by the ALSCAL program in the SPSS 11.0 software. Axes values are arbitrarily set.

This spatial representation of the authors is derived from the same correlation matrix used in the cluster analysis. The authors are each assigned an XY coordinate and plotted on a Euclidean plane. As explained in the methodology chapter, those objects further apart on the map as said to be dissimilar and those close together are said to be similar. The actual table of coordinates can be found in Appendix B. The stress value of .13099 indicates a good fit (stress levels are measured 0 to 1, with small values such as 0 to .15 indicating a good fit). The R Square (RSQ) of .91565 also indicates a good fit so I am reasonably confident of the validity of the map. Scatterplots, found in Appendix A, are relatively smooth with some gaps, but generally indicating a good relationship between the proximity and distance data.

In most all ACA studies, such as those cited in previous chapters, the resulting

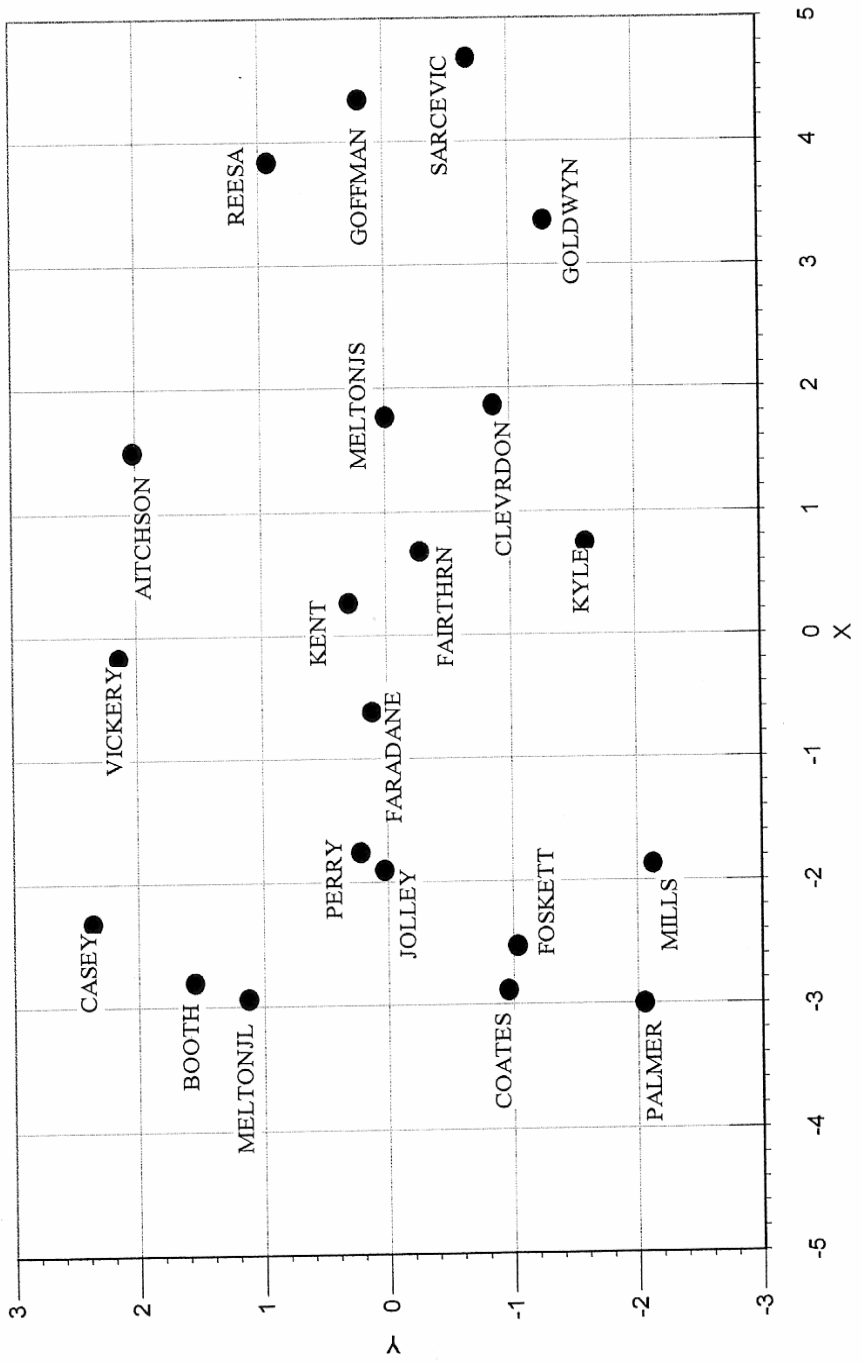



**Figure 4.2 Author map. Stress value = .13099 RSQ = .91565**

author map is enhanced by drawing circles around the clusters of authors as defined by the cluster analysis. By doing this, the "schools" of thought within the research areas in question can begin to be more assuredly identified and their location with respect to each other can be seen. This is also a good way to further analyze the proximities of the authors, as discussed above in the cluster analysis, within groups and across groups. As for the position of the authors with respect to the map's axes, the factor analysis will aid in identifying those axes. Figure 4.3 shows the same map as in figure 4.2, but with the author groups emphasized.

From left to right, one can see the four major clusters starting with cluster one in the bottom left quadrant, cluster two in the top left quadrant, cluster four spanning the two right quadrants, and cluster three occupying the central and upper portion of the map.

Cluster one contains those authors concerned with classification—its theory, education, how it pertains to the field of knowledge in general and specific subject areas, and even in its historical sense. It is a well-defined cluster, i.e., set apart on the map from the other clusters. These men—Foskett, Coates, Mills, and Palmer—I call, in the context of this study, the "grand classificationists" because they were, and still are, concerned with the *nuts and bolts* of researching and constructing traditional classification schemes and with teaching those techniques to librarians.[3] And, they embodied the "holistic" view of knowledge, as described in chapter one, in which they analyzed documents with the idea of classifying, or placing, them into an overall structure of knowledge.

Cluster two contains those authors concerned with what was then call "machine translation" or the procedures for mechanized translation and retrieval of mainly scientific literature. Melton (J.L.), for instance, wrote a number of papers on the

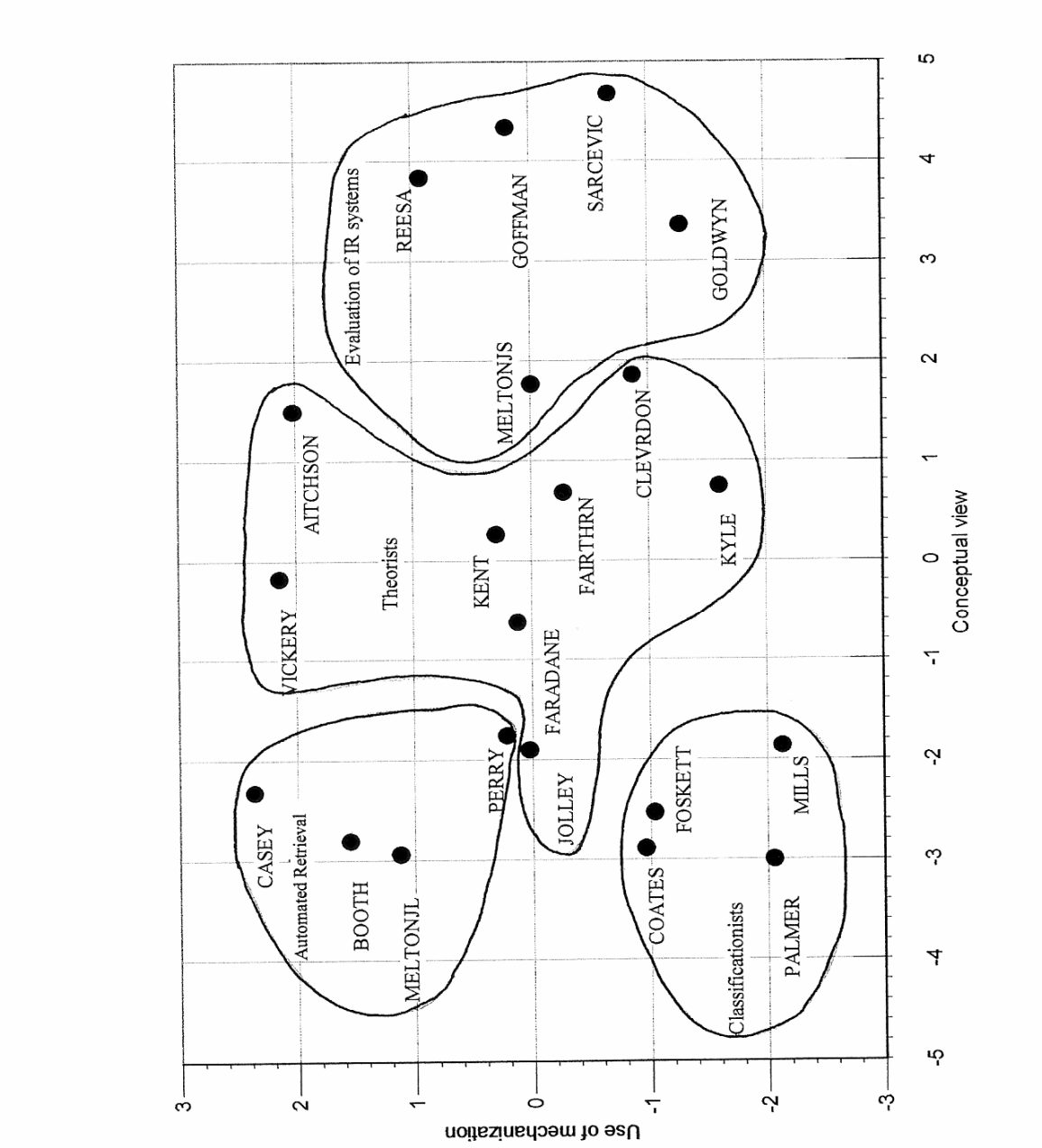



**Figure 4.3 Author map with lines drawn indicating clusters of authors.**

processed involved in the creation of the semantic code coming out of the CDCR, as well as addressing lexicographical, morphological, and syntactic issues associated with machine translation. Perry, of course, was a central force behind the early work being done at the CDCR with the punch card machines and the Rapid Selector machine used in the abstracting of this scientific literature. Casey, along with Perry, edited a much cited book on punched card applications. Booth, the only British member of this group, and overall, an outside member of LIS, was well-known for his work in machine translations, numerical methods, and automatic digital calculators. I use the label "automated retrieval, document analysis" which is a description used in White & Griffith's (1981) analysis of information science from the years 1972-1979.

Cluster three occupies the center of the map and it is here the most interesting story will be told. There are actually three sub-clusters within this one large group, according to the dendrogram in figure 4.1: Cleverdon/Kyle, Aitchison/Vickery, and Fairthorne/Farradane/Kent/Jolley. Cleverdon was at the heart of movement to evaluate information retrieval systems starting in the late 1950s and he developed the classic evaluation tools 'recall' and 'precision' which are still, even today, the bedrock measurements in IR systems evaluations. His work on the Cranfield Research Project influenced documentation research on both sides of the Atlantic and so his location on the map is close to the origin. Kyle's work centered around classification theory and construction, librarianship in terms of education, professionalism, and internationalism, and she even wrote on provision in the "intellectual welfare state" (Kyle, 1956). Most importantly, she was a prolific questioner when it came to the progress of documentation

and classification research of the time, as well as writing several reviews of the Cranfield Project reports.

Aitchison worked with Cleverdon on the Cranfield Project, but also worked on the development and analysis of a specialized faceted classification scheme, called a theasaurofacet, for the Whetstone Library at the English Electric Company. As such her work spanned both classification development and information retrieval system evaluation, but it must be noted that only work during the years of 1961 to 1963 were cited and as such she is the least influential member of the cluster.

Vickery was, and is, perhaps the most published of this group. He was the wearer of many hats, so to speak, because his work delved into the theory, the application, the traditional and the revolutionary aspects of both classification and information retrieval. Foskett (1988) wrote that Vickery had "…already established an international as well as a national reputation for practical sense based on a high level of theoretical understanding of communication in science and the factors which influence and control information storage and retrieval" (p. 203).

Fairthorne, like Vickery, wrote on a multiple of subjects such as information retrieval performance and evaluation, the mathematical aspects of classification, retrieval languages, classification, the theory of communication, and the theory of information. He was generally thought to "keep a close skeptical watch over the information scene for more than twenty years" (Brookes, 1974, p. 139). He was also wrote on the then growing division between "the so-called traditional library schools and the less conventional library methods. It is not a proper division at all, and it is doing a great deal of harm" (Fairthorne, 1962, p. 283). As well, he chided British researchers for lagging behind the

achievements of the American researchers, of which he became familiar after having spent a year and a half at the CDCR as a visiting professor.

Farradane's research on classification and relational indexing is the most cited of his works during this study's time span, including a popular piece entitled "The Psychology of Classification" in 1955. His view was that "classification is a theory of the structure of knowledge, i.e., of relations between different parts of knowledge" and that grouping, or classifying, these parts was an epistemological problem (Farradane, 1952, p. 74). His approach to this problem was view "true knowledge was scientific knowledge" and so studied scientific texts for these relations. He identified nine main relations, or operators.[4] Brookes (1986) remarked that Farradane "focused on meaning and sought the basis of this approach to relational analysis in psychological principles…that meaning lies in the relations experienced between concepts rather than in the language in which they happened to be expressed" (p. 16). Farradane also wrote on automation and scientific communication.

Jolley, too, looked for patterns in knowledge, but with a mathematical eye. In particular, he made use of set theory and attempted to factor it into what he often termed "information handling." His research spanned from representing data in data fields to the hardware needed to accomplish this—hence, several of his articles deal with punch card manipulations. He also wrote on the mechanics of co-ordinate indexing.

Kent rounds out this large cluster of authors. He is the only American and the only CDCR members included amongst the dominating British CRG members and the truly interesting aspect of this is that he was not a theorist. He has been described as having "…emphasized practical application besides the theoretical explanations of

information storage and retrieval problems" and played a "pioneering role in the field of mechanized storage and retrieval (Puri & Arora, 1988, p. 24). However, I would add that his work did seem to provide basis for others' theoretical approaches. Many of his publications focused on reporting the progress of the research being conducted at the CDCR as well as being very concerned with the issue of documentation versus traditional librarianship, which is a strong tie to the others in the cluster. He and James Perry edited or wrote numerous books and articles detailing the Center's abstracting and semantic factoring research. Later, after leaving the CDCR, he work focused on the operational aspects of centralized information services and its impact on librarianship.

The origin of the map is set by this cluster of authors and can be considered as highly representative of the two research areas in question and as such it is difficult to pinpoint exactly the character of the cluster. Farradane, Jolley, Cleverdon, Kent, and Fairthorne, with whom many of the other authors were correlated despite being clustered differently, are the most central on the map. There is a high degree of both the theoretical and the practical, but the most common factor among the central figures is their individual diversity of research.[5] They all address the organization and retrieval of information, but differ as to the methods, particularly when it comes to mechanized retrieval. Scientific backgrounds seems to play a role, but then Kyle and Aitchison were not so scientifically based. Kyle focused on the social sciences and Aitchison delved into linguistics. There is also an element of social awareness of the consequences of the direction of mechanized information retrieval and a two in particular—Kyle and Fairthorne—were very vocal with their cautions. Overall, this cluster represents a complicated crossroad of both classification research and information retrieval research

as it was experiencing a developmental explosion in the 1950s and 60s. And, it is equally important to understand that the map itself is a citation image; a picture of how these authors' works are perceived by those citing them. Obviously, their diversity of research made an impact on the citers, who seemed to thirst for both the theoretical and practical, the philosophical and the pragmatic. As such, the cluster can conceivably be characterized as composed of the major theoreticians of the time.

The fourth cluster occupies the right middle portion of the map and is composed of CDCR members only and is characterized as retrievalists evaluating IR systems using a systems-oriented approach. Retrievalists in the sense that their work represents the amalgamation and continuation of the ideas and research coming from the other three clusters and also because it can conceivable be seen as the start of the cognitive approach in IR research that dominated the 1970s and 80s. They also, along with cluster two, embodied the atomistic view as discussed in chapter one—the breaking down of a document into parts in order to look for the answers to user's queries.

Saracevic and Rees published a series of reports, papers, and procedural manuals in the 1960s detailing the work being conducted at CDCR in what was called the Comparative Systems Laboratory (CSL). CSL was a place intended for "comparative experiments…conducted to assess the influence of origin of inputs (title, abstract, full text), indexing languages, coding and format of output on the effectiveness of information retrieval systems" (Saracevic, 1964, p. 427). Specifically, they employed a 'systems approach' in which the goal was to "indicate which particular assembly of system components will provide the optimum performance (p. 427). Here again we can see the evaluation of IR systems at CDCR, but now during a time after Perry and Kent

had moved on to other institutions. Rees (1963) gave an address to Aslib in London, where several CRG members attended, detailing the research being conducted at CDCR. He reported that "we are now in a more interesting phase of activity—the search for theoretical principles which alone can give meaning to information retrieval" (p. 356). He advocated the Center's use of the systems-oriented approach—"a cycle consisting of statement of problem, construction of model design of experiments, testing of model, re-statement of problem on basis of experimental results, modification of model, further experiments, and so on" (p. 356). Finally, he stated that research in documentation was to improve the efficiency of librarianship and that in this "lies the clue to our thinking in relation to the computer" (p.361). In the sense, both Saracevic and Rees wrote much on the education of library professionals as well as the new direction of research being conducted at the CDCR. Melton (J.S.) and Goffman, too, were involved in the new research. Melton was the assistant director of the Center and wrote on the comparison of the colon classification scheme with the IR work being conducted at the Center, some on the structure of index languages, linguistics, as well as working on a NSF grant that investigated "achieving fully automatic processing of previously generated conventional abstracts for information retrieval" (Rees, 1963, p. 360). Goffman, who eventually became the director of the CDCR towards the end of its existence, both wrote and co-wrote an intriguing set of articles on the use of epidemic theory in researching scientific communication in terms of the "transmission and development of ideas within a population" (Goffman & Newill, 1964, p. 225). He also explored mathematical approaches to the spread of ideas, bibliometrics, methodologies for evaluation of IR systems, and logic. The fifth member of this group, A.J. Goldwyn, was also an associate

director of the Center. His work encompassed the problems of information in general and the design of information systems, as well as focusing on medical information activities that were part of the shift in focus of the CDCR in the early 1960s from scientific literature to the literature of the health and social sciences.

**Factor Analysis**

SPSS 11.0 yielded factor analysis results that fairly aligned with the cluster and multidimensional scaling results. Table 4.5 shows author factor loadings for four factors. Setting the eigenvalue to a minimum of 1, the correlation matrix data was reduced to four factors that are the columns in table 4.5 below. White and Griffith (1982) wrote that "…author citations are themselves interpretable as a kind of subject indexing" (p. 262). With this in mind the factors will be labeled with my interpretations of their main subject matter as indicated by the authors included on each factor, as well as taking into account the cluster characterization above.

The factors are listed in order and show the authors and their loadings on each one. Subscripts next to names indicate the an author's number of appearance on factors. All twenty-one authors load on at least one factor; nine load on two factors, one loads on three factors, and three (Farradane, Kent, and Fairthorne) load on all four factors.

McCain (1990) stipulates that "only authors with loadings greater than +/- .7 are likely to be useful in interpreting the factor, and only loadings above +/- .4 or +/- .5 are likely to be reported" (p. 440). I have listed all the factor loadings in order to show the entire factor, but I have set the cut-off point right at +/- .7. I will also discuss loadings at +/- .5 in order to show some of the cross-boundary groupings of authors.

**Table 4.5 Author Factor loadings. Display taken from structure matrix.**
**Only scores at +/-.7 were used in determining subject specialties. Decimals have been omitted.**

**Factor 1**

Classificationists; specialists

| | |
|---|---|
| Palmer | **95** |
| Coates[1] | **93** |
| Mills | **93** |
| Foskett[1] | **91** |
| Farradane[1] | **70** |
| Jolley[1] | 67 |
| Perry[1] | 58 |
| Fairthorne[1] | 49 |
| MeltonJL[1] | 49 |
| Kent[1] | 47 |
| Kyle[1] | 45 |

**Factor 2**

Evaluation of IR systems

| | |
|---|---|
| Goldwyn | **96** |
| MeltonJS[1] | **87** |
| Saracevic | **84** |
| Cleverdon | **82** |
| Goffman[1] | **78** |
| Rees | **73** |
| Kyle[2] | 68 |
| Fairthorne | 67 |
| Kent[2] | 57 |
| Aitchison[1] | 46 |
| Farradane[2] | 43 |

**Factor 3**

Automated retrieval; classification

| | |
|---|---|
| MeltonJL[2] | **95** |
| Booth | **87** |
| Casey | **85** |
| Perry[2] | **83** |
| Farradane[3] | 63 |
| Kent[3] | 60 |
| Foskett[2] | 56 |
| Coates[2] | 55 |
| Jolley[2] | 46 |
| Fairthorne[3] | 40 |

**Factor 4**

Theorists

| | |
|---|---|
| Aitchison[2] | **93** |
| Vickery[1] | **83** |
| Kent[4] | **75** |
| Fairthorne[4] | **74** |
| Jolley[3] | **73** |
| Farradane[4] | **70** |
| MeltonJS[2] | 57 |
| Goffman[2] | 52 |

subscripts: 1=first appearance, 2=second appearance, 3=third appearance, 4=fourth appearance

As stated above, the factor analysis, with loadings at +/- .7 confirms the cluster and multidimensional findings. However, at +/- .5 there are some interesting author

crossovers. The scree test, found in Appendix A, indicates a factor stopping point between 3 and 4.

Authors loading highest in factor one include all the CRG classificationists from cluster one, two authors from cluster three, and one from cluster two. These cluster two and three authors are Farradane and Jolley, also CRG members, and interestingly enough, Perry from the CDCR. I believe several reasons for Perry's inclusion are his prolific publication and domination of early documentation research at the same time that the cluster one author's were publishing. They are all also linked by being cited together in the top three or four citing publications. I feel it important to point out that these were authors who specialized in their respective endeavors, whether it be the construction of classification schemes or of machine translators.

Factor two authors comes from clusters three and four are clearly information retrieval evaluators. Goldwyn, Melton(J.S.), Saracevic, Goffman, and Rees represent the entire fourth cluster, and Cleverdon is from cluster three. His inclusion here is not surprising given the Cranfield Projects' overall influence on the methodologies developed to evaluate IR systems on both sides of the Atlantic. Kyle, Fairthorne, and Kent load at +/- .5, but it is not surprising to find them here. Kent, obviously, being a founding CDCR member and Fairthorne because he spent time at the CDCR as a visiting professor. Kyle, through many iterations of the statistical analysis of the cocitation data, is always in close proximity with Cleverdon. As discussed above, she was a keen observer of the progress of IR and wrote several reviews of the published reports of the Cranfield Project.

Factor three is dominated by cluster two members Melton(J.L.), Booth, Casey, and Perry, all loading above the +/- .7 cutoff. Loading slightly below that are Farradane,

and Kent from cluster two. Again, not surprising given Farradane's central position on the cluster map and Kent's close association with Perry and the CDCR. What is interesting is the inclusion of Foskett and Coates from cluster one, but this can be seen as the reciprocal of Perry's loading on factor one. Foskett and Coates, although classificationists in the purest sense, were an important part of the CRG's drive to improve faceted classification's role in the emerging information retrieval field. As discussed in previous chapters, the CRG rejected the current general classification systems and made the claim early on that faceted classification should be the basis of all information retrieval. In that sense, it is easy to understand their inclusion with early specialists in automated retrieval. The other important consideration is Farradane's research into concept operators and isolates in relational indexing which can be linked to Perry and Kent's work on semantic factor for their punch card abstracting service. Farradane (1957) even commented on American documentalists' "rejection of classification for information retrieval, in favour, initially, of the correlation of concepts by mechanical methods" (p. 66). This seems to represent a recognition by citers of the importance of these two areas of research.

Factor four is clearly dominated by the cluster of theorists, although it is interesting that Aitchison has the highest loading on that factor. Her work on the thesaurofacet format in the early years of the CRG is obviously of importance as her co-citations tend to fall in the most highly citing papers in this study. Kent's appearance here merits a future, more intensive, review of his work. Melton(J.S.) and Goffman's work in the later years of the CDCR produced some interesting theoretical work and so it is not surprising to see them included in this group. Goffman and Vickery, having both

done extensive research on scientific communication, have been identified as bibliometricians in White & Griffith's (1981) study of information science. I find it interesting that Jessica Melton is included here and in factor two, but not in factor one. Her 1960 technical report comparing colon classification—this system came from S.R. Ranganathan who heavily influenced the CRG—and the work at the CDCR was highly regarded, but appears not to have been seen by many outside of the United States.

The factor analysis also produces a component correlation matrix which serves to illustrate how the factors as a whole interrelate with one another. The actual matrix can be found in Appendix A, but figure 4.4 illustrates the correlations more clearly.

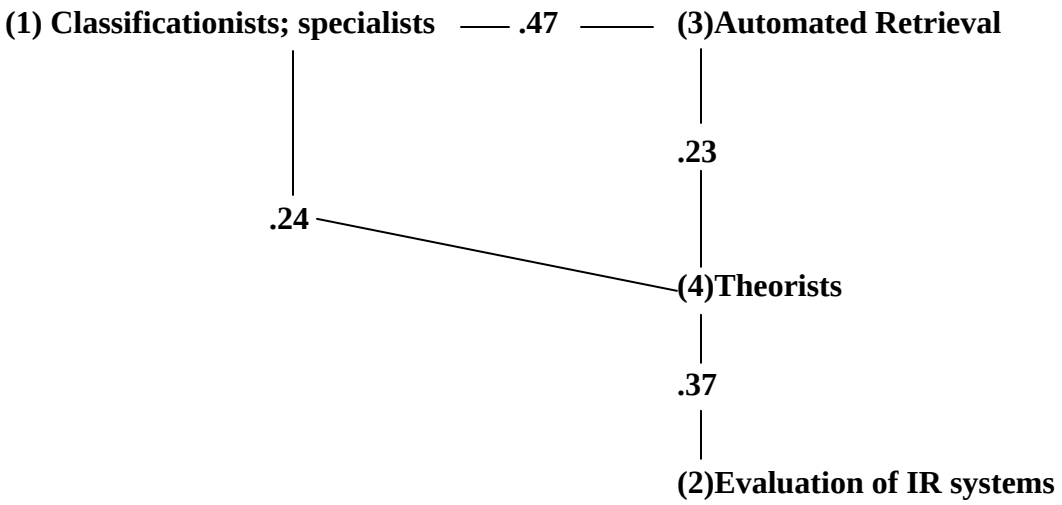



**Figure 4.4 Component correlations with strength of relationship indicated.**

There is a strong link between factors one and three (.47) and between factors four and two (.37). Factor four has moderate correlations to factor one (.24) and factor three (. 23). Lastly, factor two does not seem to correlate with factors one and three. McCain (1990) points out that intercorrelation among factors may "reveal subject-related linkages *above the author* level" and that uncorrelated ('independent") factors "may not yet have been linked by citers—an important insight" (p. 440). Given this, this factor analysis suggests that the information retrieval work coming out of the CDCR in the 1960s, having moved beyond the earlier automated work dominated by Perry and Kent and influenced by the developmental theoretical work coming from cluster three authors (as indicated by the correlation between factor two and four) had distanced itself greatly, in the eyes of those citing these works, from the classification research coming from the CRG members represented in cluster one. Figure 4.5 below shows the final author map with both clusters and factors emphasized and with the horizontal and vertical axes defined. The x-axis is considered a measure of the conceptual view of the authors, moving from a holistic view on the left to an atomistic view, as discussed in chapter one. The Y-axis indicates use of mechanization, starting with non-mechanical at the bottom and moving upward in use. I must note that this is a two-dimensional map and as such there are limitations on the number of dimensions, i.e., the two axes are only one layer of dimensionality. Further research on these clusters would reveal deeper dimensions of the map.

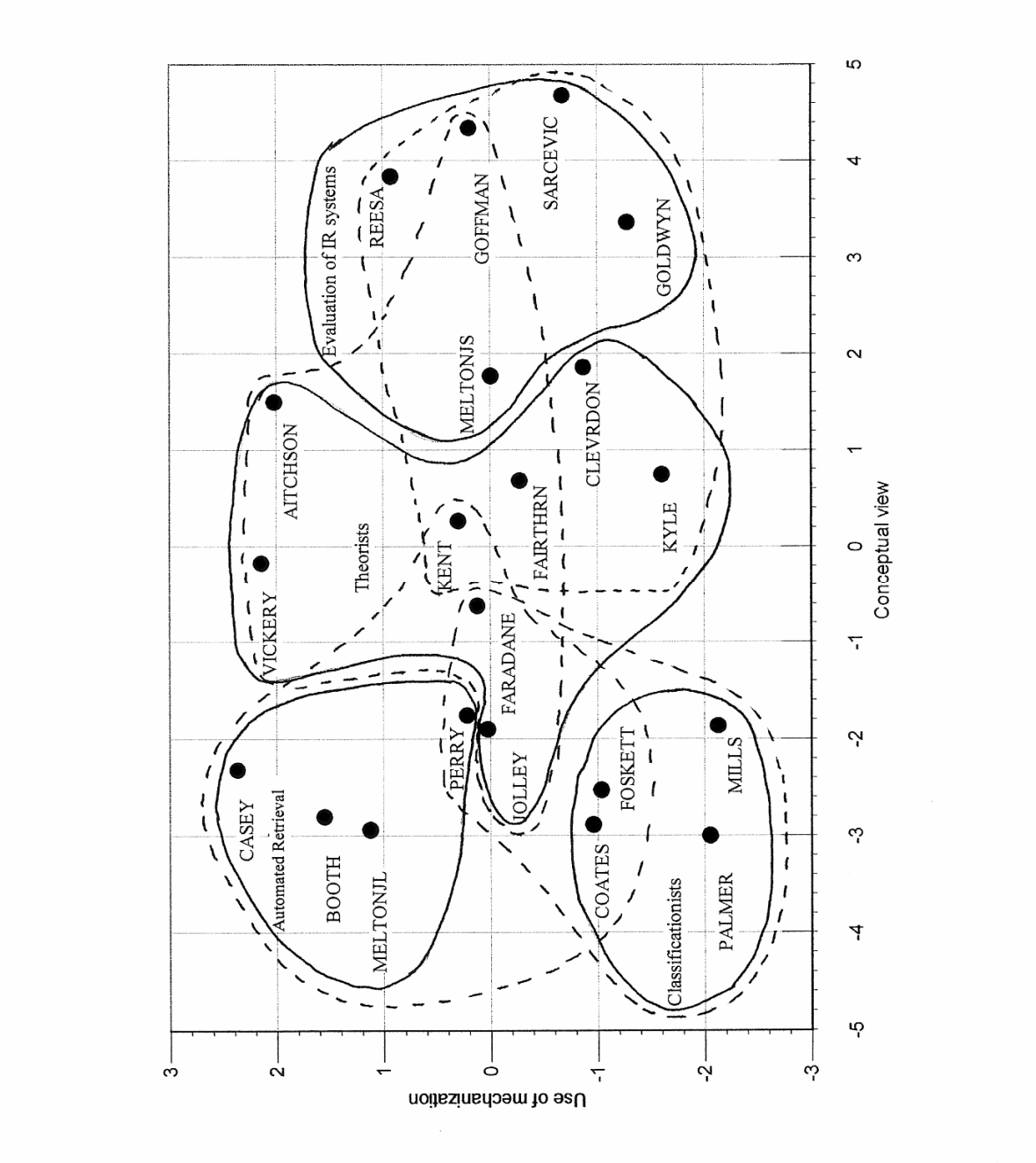



**Figure 4. Final author map with clusters, factors, and axes defined.**

## Interpretation and Validation

### Inspection

The author cluster map displays an interesting mix of persons, consisting of library practitioners, scientists, chemists, mathematicians turned documentalists or classificationists, library school teachers, social scientists and administrators.[6] I have attempted to provide as much background information as possible for each author included in this study, but have so far focused mainly on the content of their work. I have also made judgments, based on the statistical evidence, on the "make-up" of each of the four clusters of authors and have labeled them accordingly. I will now attempt to define the social types, worldviews, information behavior, and social norms of each group as a method of interpreting the small worlds of each.

### Application of concepts taken from the theory of normative behavior

**Worldview**. The author clusters tend to consist of either only CDCR members or only CRG members, with only Kent being included with CRG members in cluster three. As discussed in chapter one, library classification research was greatly influenced by the idea of a natural order, or structure, of knowledge and this in turn related to documents in that the goal was to try to 'fit', or classify, documents into that order or structure. Machine-based document retrieval focused on techniques for sorting documents and while this *may* have been related to an overall knowledge structure, those creating these document retrieval systems were not really interested in that overall structure. Rather, they preferred to view document retrieval as the intersection of a person's question with a document or documents without benefit of an overall structure of knowledge. Cluster one

seems to hold the first worldview, that a belief in overall structure of knowledge was the best approach to organizing and providing access to information. Cluster four, those who evaluated information retrieval systems, seem to hold the second worldview, that breaking down documents, and to a large extent the information systems themselves, into separate pieces would facilitate more efficient information retrieval.

Cluster two, associated with early automated retrieval of information, actually did base their early punch card machines on abstracts of the documents, thereby keeping some semblance of the wholeness of a document. They also investigated and developed a process of semantic factoring which implied some overall structure, but not in the same sense of traditional library classification systems.

Cluster three's worldview is difficult to define because of the diversity of the make-up of the authors. As theorists in general, their work seemed to be greatly experimental and implies that there was much shifting of worldviews, almost like trying to find a comfortable pair of shoes to wear for a long walk. If we read the author map from left to right we can see the progression from holistic to atomistic and I believe this comes from the disruption the computerization movement had on the traditional views and the division it created between librarians and documentalists. This in turn leads to a discussion on social types.

**Social Types.** Cluster one generally consists of British practicing librarians and library educators. I will take the cluster characterization of 'classificationists' as an indicator of their social type as well. Cluster two consists of scientists turned documentalists, Perry and Kent, and Booth, who was an Interdisciplinary Professor of Autonetics. Casey is enigmatical, but at the time he edited a book on punch cards with

Perry he was based at the W.A. Sheaffer Pen Company in Fort Madison, Iowa. The best description of these men is that they saw themselves as occupying the role of machine translators or even machine retrievalists that would head up the information retrieval movement. Cluster four, however, consists of information retrievalists in the modern sense of the word, i.e., as it is used presently in the field of LIS.

Again, cluster three is hard to define in terms of social types. Vickery was the librarian at the ICI Butterwick Research Laboratories in England and it was his study of Bradford's law of scattering for "practical use in an industrial library" that led him into the arena of library services (Foskett, 1988, p. 199). He work over the years encompassed most "every aspect of library and information work" (p.202). He was one of the founding members of the CRG and even though he drifted away from it sometime in the 1960s his work in general was, and still is, a great influence upon the whole of library and information science. He is generally thought of as a "theoretician," but his "career demonstrated a perhaps unusual capacity to contribute both to the fundamental understanding of the nature of information science and to the development of practical services as a result of such insights" (East, 1993, p. 855). Barbara Kyle, as discussed earlier in the chapter two literature review, was an influential member of the CRG, but that was cut short by her premature death in the mid-1960s[7]. Wilson (1965) describes her background in public libraries and as the Librarian of the Royal Institute of International Affairs, as well as "a brief incursion into the world of publishing as Assistant Director of the National Book League" (p. 227). She was also involved with the Association of Special Libraries and Information Bureaux (Aslib) as both a Research Librarian and as co-editor of the *Journal of Documentation*. At various times in her

career “she passed from the straightforward work in libraries through special library work associated with the social sciences to… an abiding interest in classification and indexing” (p. 230). Before her death she had started work on the Kyle Classification, a unique classification system intended for the whole of the social sciences.

Cleverdon is perhaps best known for his work on the Cranfield Research project, beginning in 1957 and ending in 1967, which involved comparison of “four index languages: UDC, alphabetical subject catalogue, Uniterms, and a special faceted classification” (Lancaster, 1993, p. 212). At the time of the project he was Librarian at the College of Aeronautics, Cranfield, England, hence it is called by the more familiar name “Cranfield Project.” His work in the ensuing years involved important work on information retrieval and his influence was “widely recognized in North America, Europe, and elsewhere” (p. 213). Several of the authors in this study worked with him on the Project, including Aitchison, Vickery and Fairthorne from the CRG. Even more interesting is his interaction with the researchers at the CDCR who took an active interest in his information retrieval evaluation work. Indeed, Cleverdon remarked once on the “influence on him of James Perry’s editorial comment advocating ‘cautious and searching evaluation of experimental results’ in *American Documentation* in 1955” (Keen, 1998, p. 267). Cleverdon visited the CDCR sometime in 1962 and a friendly debate ensued between himself and Allen Kent on the value of using computers in information retrieval. Cleverdon remarked that “…he could see no real use for computers in information retrieval until someone builds a machine that can think” (“Briton devalues”, 1962). Kent replied that “…it has usually been found that you can’t mirror every human task with a machine” but that further research can reduce the amount

of irrelevant data (Abraham, 1962). He did agree that information retrieval still had to begin and end with a human librarian.

CRG member Robert Fairthorne was also associated with the CDCR in the 1960s. He spent a year and a half doing research at the Center in 1961 and 1962 as a Visiting Research Professor and was also appointed to its Advisory Board. Fairthorne started his career as a mathematician as the Royal Aircraft Establishment in Farnborough, England, before the outbreak of World War II. It was during WWII, however, that he created "a complete computational laboratory…used to solve a diversity of problems, including some in general statistics as well as those in strictly aeronautical subjects" (Walker, 1974, p. 128). It was here as well that he used a punched-card machine called the 'Hollerith' to work on classification and allied problems. In the late 1940s he was the Secretary of the British Society for International Bibliography and "his output of papers on information storage and retrieval started growing" indicating his a shift in his interests from aeronautics to documentation (Coblans, 1974, p. 131). Brookes (1974) wrote that Fairthorne's primary contribution to information science was "to define its scope, to clarify its terminology, and to establish its fundamental principles" (p. 139).

Jason Farradane was another chemist-turned-documentalist who contributed much the CRG and to the LIS field in general with papers focusing on abstracting and indexing issues, as well as classification theory and the importance of library education. In 1958 he co-founded the Institute of Information Scientists and was Director of the Centre for Information at The City University in London. In the mid-1960s he was the editor-in-chief of the journal *Information Storage and Retrieval*. Justice (2001) found Farradane to have an interesting history in that he changed his last name from Lewkowitsch to

Farradane by combining the names *Faraday* and *Haldane*, two names that "invoked the primacy of the scientific, rationalistic approach to all social activities" (p. 6).[8]

Jolley was a mathematician, Kent was another chemist-turned-documentalist, and Aitchison, who worked at the English Electric Company, had an abiding interest in linguistics and thesaurus construction. The role each of these researchers saw themselves as playing in the LIS field during the eighteen year time span is difficult to define. I believe researchers such as Kyle and Fairthorne say themselves as "watchdogs," in some ways. As I stated earlier in chapter three, they both cautioned the field in general when it came to overzealous use of computers. Kyle (1963), writing a guest editorial for *American Documentation*, asked of American documentalists:

> What this European observer wants to know is: how far does the work done with all this hardware and all these dollars influence the general run of documentation work throughout the country? (p. 93)

Overall, the social types of all four clusters, at the time in question, seemed to rest either on the complete belief in the use of computerized organization and retrieval of information or on the skeptical view of these machines and a great reluctance to accept their usefulness without first establishing foundational principles governing their use and development. This in turn governed how they each saw their role in the field in general. I don't see evidence of either the CRG members of the CDCR members as aligning themselves completely with "librarians" or with "documentalists" as it pertains to the division that was perceived between these two as discussed by Farkas-Conn and Bowles in chapter one. They were, for the most part, outside observers of this division and some,

Kent, Rees, Goldwyn, Fairthorne, for example, believed their research could help to mend the division.

**Information behaviors and Social norms.** This study employs a co-citation analysis in which data is gathered on who is citing these two groups of researchers and not data on whom the two group of researchers are citing. Thus, defining information behaviors is limited and defining social norms must be set aside for future research; I had asked in chapter one if the citing behaviors of the CRG reflected the need to move away from the older ideals of classification research, i.e., were they citing more IR authors or were they remaining within their own small world of accepted classification writings? I can not answer this with this present study because I have not analyzed who either the CRG or the CDCR were citing in their works. However, I can make some conclusions as whether or not this "move" by the CRG was acknowledged by those outside of their small world, i.e., by their peers. Thus, I can attempt to define their information behavior by how they were perceived by those outside their small worlds. For example, the classificationists in cluster one were very set apart from the other clusters on the author map and according to the factor correlation (see figure 4.4) they were very weakly connected to the information system evaluators in cluster four who were spearheading the new IR movement. Cluster one is, however, strongly correlated with clusters two and three so I can conclude their intent to move from older, more traditional classification was acknowledged, but not to any great extent, by the up and coming IR researchers. Those CRG members contained in cluster two more often than not crossed the line into IR research in various ways, Vickery and Cleverdon being the best examples of this behavior. Hence, there was a degree of overlapping citing behavior, but only with those

who chose to research and publish on IR developments and not align themselves solely with classification research. This seems to reflect the perception that traditional classification research on the whole was not useful in IR studies at the time.

**Conclusions on the four concepts.** There are some outside influences that need to be acknowledged as having effected the small worlds of these two groups of researchers. I have already discussed the division between librarians and documentalists as it pertains to the two groups of researchers. There is also the issue of the differences in American and British science. Science research in the United States, specifically in the area of documentation during the time span specified in this study, was highly funded by both government and corporate sponsors. Jesse Shera, who started the CDCR and put it into the hands of Perry and Kent, was in close contact with Senator Hubert Humphrey.[9] In fact, Perry and Kent testified at congressional hearings and encouraged the creation of such agencies as the National Referral Center for Science and Technology and the Science Information Exchange (Curtis-Wright, 1988, p. 26). There was a great push for developments in documentation resulting out of the post-WWII information explosion, and the shock of the Russians launching Sputnik. Miksa (1987) points out the rise of research teams—formally organized teams instead of isolated individuals—as being an off-shoot of the time when the "necessity of research became a factor in all fields and in American society in general" (p. 4). In this sense, the CDCR was more funded and, as such, more motivated to produce results. They were certainly more inclined to pass over traditional classification research by sometimes labeling it with the term "pigeonhole".

An the other side of the spectrum are British science practices which seemed to move at a slower, more philosophical pace. Specifically with regards to librarianship, the

Library Association, similar to the American Library Association, was an important part of the research in librarianship and for supporting library education. However, although it "encouraged research into library and bibliographic problems from its early days, it has never had the financial resources of the staff to initiate and carry out major research projects" (Haslam, 1975, p. 333). Indeed, it wasn't until 1963 that a committee was set up to investigate the problems of "postwar information explosion in the field of scientific and technical literature" and no report was issued until 1968 (p. 329). There was also the Royal Society of London, a time-honored scientific society who at the 1948 Royal Society Scientific Conference, declared that a study of classification should be made and so the CRG was established formally in 1952. However, even with that support Foskett (1962) wrote that

> during these ten years the C.R.G. has met nearly every month, and although it has never had more than about a dozen active members, its influence has grown to the point at which is causes Mortimer Taube in America to rage over its medieval scholasticism, John Metcalfe in Australia to denounce it as a plot by Ranganathan to ruin librarianship, and a British University librarian to describe it as one of the two most significant developments in British librarianship since the end of the war. (p.127)[10]

Overall, while the American CDCR was more stabilized by corporate and government support and by its own insatiable thirst for mechanization, the British CRG, set on a task to research and better library classification, seemed to be ostracized even by it own people.[11] Foskett (1962) noted that "classification is thought of by many librarians as

either a fearful complication of a very simply act, or an outmoded, almost prehistoric, method of doing a very complex mathematical task" (p. 138). Furthermore, the CRG seemed to regard the documentation work coming out of the states with a weary eye, saying that "the CRG would be hard put to ignore this [punch cards and its hardware], even if it had wanted to, which it does not" (p. 137). Overall, these outside influences had a large effect on the shaping of the social types, information behavior and worldviews of the both the CRG and the CDCR.

## Endnotes

[1] This raw data matrix is too large to contain in this dissertation but is available upon request.
[2] McCain indicated to me that a frequency analysis is used more so to ensure there are no missing values then to compare frequencies (personal conversation, 2001).
[3] I spent time with Coates, Mills, and Foskett in the summer of 2001 and the spring of 2002 and observed first hand this process of classification construction as they worked on various sections of the Bliss Bibliographic classification system during CRG meetings. In doing so, I found their present *attitude* much the same as I would read in their published works from the 50s and 60s. During these two meetings they would often stop and instruct me as to what they were doing—always the teachers.
[4] These are: concurrence, equivalence, distinctiveness, self-activity, dimensional, action, association, appurtenance, and functional dependence.
[5] I would liken it the spokes of wheel radiating out in all directions.
[6] Again, "Documentalist" here is equal to "information scientist" just as *documentation* is equal to *information retrieval* within the context of the time period in question.
[7] I have found in my research of the background for this study that many authors often wonder what the LIS field in general and classification research in particular would have been like if Kyle had not died, so much so was her influence. Wilson (1965) wrote that "the very purposes for which Aslib stands are basic to her concept of human progress: as knowledge is the key to man's advancement, so is the ordered arrangement of knowledge the key to its understanding and use" (p. 227).
[8] These refer to the Faraday Society and possible to J.S. and J.B.S. Haldane, father and son scientists (Justice, 2001, p.6).
[9] In my archival research at CWRU I found several letters between Shera and Senator Humphreys.
[10] Taube and Metcalfe were both earlier influential pioneers in librarianship.
[11] See Foskett's comment in chapter 1, page 18.

# CHAPTER 5

# CONCLUSIONS

## Restatement of the Problem

This dissertation has examined two subfields within the field of library and information science—classification research and information retrieval research—during an eighteen year time period in order to ascertain how a division between these two areas of research had occurred. I undertook a parallel case study of two groups of researchers within these two sub-fields: the Classification Research Group (CRG) and the Center for Documentation and Communication Research (CDCR). The CRG and the CDCR represent the most recognized and organized groups within the areas of classification research and information retrieval research from 1952 to approximately 1970.

The justification for the research stems from the present day perception by main information retrieval researchers that traditional classification research is simply a mechanism for placing and locating books on shelves or that it is an 'information losing' process that has no real place in the present day online environment. A brief survey of information retrieval textbooks and conference proceedings has revealed this perception and as such begs the questions as to how this perception came about, especially in light of the substantial amount and quality of classification research that has been conducted over the few last decades. In addition, developers of classification schemes used in Internet

search engines (often referred to as taxonomies or ontologies) are encountering the same problems experienced by classification researchers sixty years ago, but consistently fail to investigate or benefit from these past experiences.

## Research questions

Three research questions were put forth that would begin to address the problem as stated above. First, what was the initial relationship between the Classification Research Group and the Center for Documentation and Communication Research and did this relationship change between 1952 and 1970? Based on review of historical literature on both groups I have found that the CRG and CDCR members had professional knowledge of each other and, for the most part, they regarded each other's work as important to the LIS field in general. They shared a common belief that the current classification systems were inadequate for the information demands of the time and each looked hopefully on emerging computer technologies for solutions to these problems. There was some interaction at international conferences in the 1950s and early 1960s, as well as a few interspersed visitations by members of the CRG to the CDCR in Cleveland, Ohio. As the years progressed, however, there appeared to be a lessening of interaction due to changes in membership and research trends. In the mid-1960s the CRG turned from its development of specialized classification systems back to developing one general, or universal, scheme and the CDCR had re-focused its research to encompass a system-oriented approach to information retrieval system evaluation.

The second research question asked what were the conceptual similarities and differences of these two groups as evidenced in their published artifacts from 1952 to 1970? An author co-citation analysis was performed using three types of multivariate

analysis—cluster analysis, multidimensional scaling, and factor analysis—in order to shed light on how the work from these two groups was used by the LIS field overall. This was a limited analysis in that it looked only at how the two groups were being cited and not at who they were citing in their own research. As the citing of published research is considered a recognized form of scientific communication, I felt this analysis would provide a good basis for answering this second question.

Findings from the author co-citation analysis demonstrate that among the twenty-one authors identified as having sufficiently significant co-citation frequencies to use in the study four distinct clusters of authors can be identified. Two clusters contain only CDCR members (four in one, five in the other), one contains only four CRG members, and the largest cluster contains seven CRG members and one CDCR member. The cluster containing only CRG members has been characterized as classificationists. The two clusters containing only CDCR members has been characterized as automated retrievalists (4), who represent the earliest years of the CDCR, and evaluators of information retrieval systems (5), who represent the middle to later years of the CDCR.

Multidimensional scaling placed the authors on a two-dimensional spatial representation according to author proximity counts (typically called an author map). The four author groups resulting from the cluster analysis were then matched to this map and the two analyses were found to be in agreement. The proximity of the authors, derived from a correlation matrix of the authors' co-citation counts, indicates that one's position on the map can be argued to mean either intellectual closeness or intellectual distance. Those authors nearest to each other can be construed as sharing similar work and subject matters in their published artifacts. Those authors furthest from each other

can be construed as intellectually distant from each other. The two groups the most distant from each other were the classificationists and the evaluators of information retrieval systems. However, the IR system evaluators, all CDCR members, were also distant from the early CDCR members characterized as automated retrievalists. Most of the large cluster three authors were linked in some way to members in the other three clusters, which indicates the cross-boundary nature of those authors' research.

The factor analysis also coincided with the other two analyses. It produced four distinct factors on which the authors loaded in much the same rank as they were grouped in the clusters or positioned on the map. However, there were again a few notable cross-overs from the central cluster of theorists, indicating their influence on both research groups. Taken together, these three analyses demonstrate that a division did appear between the group of classification researchers and those information retrieval researchers who tended toward IR evaluation methodologies.

The third question asked how could we characterize the relationship between the small worlds of the two groups using the theory of normative behavior and its concepts of worldviews, social norms, social types, and information behavior? Firstly, I deliberately ignored social norms as I did not collect citing data from the authors included in both groups. This can be tested in a future study. Secondly, the information behavior can only be seen through the filter of those citing the authors in question. The worldview of the two groups seems to center around either a holistic or atomistic view of information organization and structure; i.e., retrieval of information can be accomplished by placing an item within an overall structure of knowledge to show it places and relationship to other items or the item in question is broken down into its collective parts and matched to

the user's query. This query is also broken down into its collective parts. These parts that are of a similar nature are matched and then the item itself is retrieved and, theoretically, a useful answer is provided to the user.

The identification of social types was more difficult as it was placed within the context of the division between librarians and documentalists that was occurring during the 1950s and 60s which in turn was an outcome of the overwhelming urge of scientists and documentalists to computerize everything. The diversity of educational backgrounds of the members of both groups also made it difficult to pinpoint types. CRG members seemed to hold themselves separate in ways from librarians in general and some regarded themselves as very distant from 'punch-card enthusiasts,' but some members did embrace the new technology and were intrigued enough by it to develop some fundamental principles for guiding its development or at least to caution others in its use. Hence, there are two clusters of CRG members. CDCR members seem to align themselves with either complete machine translation or with developing systems that aided in information retrieval in library settings and with evaluating those systems. Hence, there are two distinct groups of CDCR members.

Lastly, the external factors of differences in scientific methods and pace, as well as the always present money and funding situation seem to create two somewhat distinct small worlds of classification researchers and information retrieval researchers during the eighteen year period under study.

## Final Conclusions

As stated in chapter one the significance of this study lies in its attempt to better understand why two areas of research in the field of library and information science diverged when the ultimate goals of each—effective organization and retrieval of knowledge—are so clearly related. This was not an exhaustive study of the whole of either classification research or of information retrieval research. By examining two sub-groups with the library and information science field I attempted to demonstrate with evidence the very beginning of an explanation for how it occurred and in doing so have come across a small portion of the "why" as well. The final conclusion must be to continue examining the hints and to keep generating the guesses.

# CHAPTER 6

# FUTURE RESEARCH

## Bibliometric Coupling

There are many avenues for future research stemming from this study. The most logical one will be to perform the reciprocal of the author co-citation analysis—a bibliometric coupling analysis. Bibliometric coupling is the study of the co-citing behavior of two or more authors, i.e., who do they cite and what does this reveal about the degree of communication. I feel this is best next step in attempting to solve the problem I have investigated in this study. If I can reveal who the CRG and CDCR cited in common then I can make further conclusions about the division that occurred between classification research and information retrieval research. In addition, by analyzing who the two groups cite in common I can create a better picture of the social norms and information behavior. Specifically, I can attempt to answer the questions I asked in chapter one pertaining to these two concepts: Were there "boundaries" where citing certain author's works were concerned and did the members of either group feel inclined or disinclined to cross those boundaries? Do the citing behaviors of the CRG reflect the need to move away from the older ideals of classification research, i.e., were they citing more IR authors or were they remaining within their own small world of accepted classification writings?

## Content Analysis of Author Oeuvres

I have attempted in some small ways to analysis the content of the published works of the CRG and the CDCR, but not to the extent of a comprehensive content analysis. As stated in chapter three, White and Griffith (1981) write that

> *Oeuvres*, we conclude, are inherently ambiguous objects of study. To uncover why they appear on the [author] map as they do, one must make a detailed examination of the writings in them and of how and why they are cocited with writings by other authors. This, of course, involves an analysis at the level of specific documents, which is beyond our present scope, but may be possible in other contexts. (p. 166)

I feel an extensive analysis of the works of the CRG and CDCR identified by the author co-citation analysis would greatly lend itself to better understanding the division in research by getting down to the actual conceptual level of ideas in the works and to see how they differed or came together. Furthermore, would it be possible take the results of the level of analysis and demonstrate some of the shortcomings of present day research in both areas of research?

## Holistic versus Atomistic

The idea of holistic and atomistic views of information handling is intriguing in and of itself. A deeper exploration into the fundamental basis for both these concepts as it pertains to information organization may eventually lead to the development of a theory of on the differences. For example, examining the intersection of a person's question with a collection of documents organized in a structure of knowledge to an intersection of a person's question with a collection of documents that are not organized

with any structure of knowledge—how are library and information scientists dealing with these two approaches in their research-- are they even addressing it? Are they even aware of it? This leads to an tangent question: implicit versus explicit structure in information organization and retrieval systems.

## Implicit versus Explicit Structure

In looking at structured approaches to organization there can be either implicit structure or explicit structure; structure that is internal and can't be seen and structure that is external and is easily discernable. *Structured* as a concept alone is tantamount to imposing some sort of form onto information, or knowledge, within the confines of an information system by a series of techniques and using many different tools. It has long been stated that users need structure to assist in searching. For instance, with subject heading cross references one can narrow or broaden one's search. And, with Library of Congress Classification or Dewey Decimal Classification one can narrow or broaden a search and find related subjects all collocated nearby—the subject one zero's in on. This is somewhat vague because it doesn't differentiate between implicit structure and explicit structure. We can restate the assumption in another way: "User's need implicit structure to assist in searching." Obviously they do because all information retrieval needs, or will, resort to some sort of structure quite naturally.

We can restate the assumption with explicit structure: "User's need explicit structure to assist in searching." Here a problem appears. This may well have been what those statements made by groups like the CRG have meant over the years about needing structured knowledge organization. In short, library subject access people may have

assumed this, but is it true? Or, if it has some truth, can we distinguish between different situations or cases?

We can restate the assumption again: "User's need explicit structure to assist them in doing known-item searches." It can be claimed they certainly need the structure that is provided with organizing citations into descriptive fields (i.e. categories of descriptive data). But, do users need structure to do CATEGORY searches? In other words, does it help them to know that the subjects for which they search are related to broader terms, to narrower terms, or to related terms in the system? Again, perhaps.

The case of a fully structured and displayed system like LCC or DDC is more problematic. Under what conditions might a user appreciate and be able to use a full display of the various facets of any classes to which a subject is related by being part of such a class's hierarchy or by being part of the hierarchy of a related class? We don't show the user those more complicated class structures, or at least not up until the advent of the online catalog. The user never sees them, has never seen them, unless they have been introduced to the schedules of the DDC or the LCC, which some bibliographic instruction classes may well do. These complicated structures are implicit in the call numbers that users see on items on the shelves, but the numbers are not labeled with their class names. We do sometimes put up something off the top of the scheme hierarchies in some public place-- the first two summaries of the DDC or the letters of the LCC, but do users find them sensible and usable or do they think of them as stack guides and not the properties of a structured system?

The long and the short of this is the assertion (or, simply the claim) that users need explicit structure for category searches. I don’t think we really know much about it,

and in fact the user has rarely had that kind of information given to him or her in any sensible way. I think this area needs research, the findings may be surprising.

### Bibliometric analysis and classification systems

One interesting off-shoot of this study has been the question of how can bibliometric analysis of scientific work contribute to, or intersect with, the development of classification systems? For instance, is the information behavior of researchers--how, who, what they cite in their papers--reflected in the classification systems that deal with those subject areas? For instance, if I were to do a citation analysis of a specific area of scientific research and found the structure of that area of research as perceived by the citers would that structure correspond to how that area of research has been organized by the classification system?

These are just some of the directions that research stemming from this study may go. Ideally, the most important first step is to complete the citation study by performing the bibliometric coupling analysis and by embarking on an extensive content analysis of the published works of the CRG and the CDCR.

# APPENDIX

## Multivariate Analysis Syntax from SPSS 11.0

```
FREQUENCIES
  VARIABLES=aitchson booth casey clevrdon coates fairthrn faradane foskett
  goffman goldwyn jolley kent kyle meltonjl meltonjs mills palmer perry reesa
  sarcevic vickery
  /ORDER= ANALYSIS .

MULT RESPONSE groups=authors (aitchson booth casey clevrdon coates fairthrn faradane
foskett
  goffman goldwyn jolley kent kyle meltonjl meltonjs mills palmer perry reesa
  sarcevic vickery (1) )
/TABLES=authors BY authors .

FREQUENCIES
  VARIABLES=aitchson booth casey clevrdon coates fairthrn faradane foskett
  goffman goldwyn jolley kent kyle meltonjl meltonjs mills palmer perry reesa
  sarcevic vickery
  /ORDER= ANALYSIS .

PROXIMITIES  aitchson booth casey clevrdon coates fairthrn faradane foskett
  goffman goldwyn jolley kent kyle meltonjl meltonjs mills palmer perry reesa
  sarcevic vickery
  /MATRIX OUT  ('C:\WINDOWS.000\TEMP\spss4294636009\spssclus.tmp')
  /VIEW= VARIABLE
  /MEASURE= CORRELATION
  /PRINT  NONE
  /STANDARDIZE= NONE .
CLUSTER
  /MATRIX IN  ('C:\WINDOWS.000\TEMP\spss4294636009\spssclus.tmp')
  /METHOD COMPLETE
  /PRINT SCHEDULE
  /PRINT DISTANCE
  /PLOT DENDROGRAM VICICLE.
ERASE FILE= 'C:\WINDOWS.000\TEMP\spss4294636009\spssclus.tmp'.

PROXIMITIES aitchson booth casey clevrdon coates fairthrn faradane foskett
  goffman goldwyn jolley kent kyle meltonjl meltonjs mills palmer perry reesa
  sarcevic vickery
/MATRIX OUT('C:\WINDOWS.000\TEMP\spss4294636009\spssalsc.tmp')
/VIEW=VARIABLE
/MEASURE=CORRELATION
```

```
/PRINT NONE
/STANDARDIZE=NONE .
ALSCAL
  /MATRIX= IN('C:\WINDOWS.000\TEMP\spss4294636009\spssalsc.tmp')
/SHAPE=SYMMETRIC
/LEVEL=ORDINAL (SIMILAR)
  /CONDITION=MATRIX
  /MODEL=EUCLID
  /CRITERIA=CONVERGE(.001) STRESSMIN(.005) ITER(30) CUTOFF(-1) DIMENS(2,2)
  /PLOT=DEFAULT ALL .
ERASE FILE='C:\WINDOWS.000\TEMP\spss4294636009\spssalsc.tmp'.

FACTOR
  /VARIABLES aitchson booth casey clevrdon coates fairthrn faradane foskett
  goffman goldwyn jolley kent kyle meltonjl meltonjs mills palmer perry reesa
  sarcevic vickery  /MISSING LISTWISE /ANALYSIS aitchson booth casey clevrdon
  coates fairthrn faradane foskett goffman goldwyn jolley kent kyle meltonjl
  meltonjs mills palmer perry reesa sarcevic vickery
  /PRINT INITIAL EXTRACTION ROTATION
  /FORMAT SORT BLANK(.4)
  /PLOT EIGEN
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION .
```

## SPSS 11.0 Frequencies Distribution

# Frequency Table

**AITCHSON**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 159 | 87.8 | 87.8 | 87.8 |
| | 1 | 22 | 12.2 | 12.2 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**BOOTH**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 175 | 96.7 | 96.7 | 96.7 |
| | 1 | 6 | 3.3 | 3.3 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**CASEY**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 162 | 89.5 | 89.5 | 89.5 |
| | 1 | 19 | 10.5 | 10.5 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**CLEVRDON**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 113 | 62.4 | 62.4 | 62.4 |
| | 1 | 68 | 37.6 | 37.6 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**COATES**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 167 | 92.3 | 92.3 | 92.3 |
| | 1 | 14 | 7.7 | 7.7 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**FAIRTHRN**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 137 | 75.7 | 75.7 | 75.7 |
| | 1 | 44 | 24.3 | 24.3 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**FARADANE**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 155 | 85.6 | 85.6 | 85.6 |
| | 1 | 26 | 14.4 | 14.4 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**FOSKETT**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 158 | 87.3 | 87.3 | 87.3 |
| | 1 | 23 | 12.7 | 12.7 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**GOFFMAN**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 153 | 84.5 | 84.5 | 84.5 |
| | 1 | 28 | 15.5 | 15.5 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**GOLDWYN**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 166 | 91.7 | 91.7 | 91.7 |
| | 1 | 15 | 8.3 | 8.3 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**JOLLEY**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 175 | 96.7 | 96.7 | 96.7 |
| | 1 | 6 | 3.3 | 3.3 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**KENT**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 140 | 77.3 | 77.3 | 77.3 |
| | 1 | 41 | 22.7 | 22.7 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**KYLE**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 165 | 91.2 | 91.2 | 91.2 |
| | 1 | 16 | 8.8 | 8.8 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**MELTONJL**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 178 | 98.3 | 98.3 | 98.3 |
| | 1 | 3 | 1.7 | 1.7 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**MELTONJS**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 165 | 91.2 | 91.2 | 91.2 |
| | 1 | 16 | 8.8 | 8.8 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**MILLS**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 166 | 91.7 | 91.7 | 91.7 |
| | 1 | 15 | 8.3 | 8.3 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**PALMER**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 172 | 95.0 | 95.0 | 95.0 |
| | 1 | 9 | 5.0 | 5.0 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**PERRY**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 152 | 84.0 | 84.0 | 84.0 |
| | 1 | 29 | 16.0 | 16.0 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**REESA**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 142 | 78.5 | 78.5 | 78.5 |
| | 1 | 39 | 21.5 | 21.5 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**SARCEVIC**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 165 | 91.2 | 91.2 | 91.2 |
| | 1 | 16 | 8.8 | 8.8 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

**VICKERY**

| | | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|---|
| Valid | 0 | 111 | 61.3 | 61.3 | 61.3 |
| | 1 | 70 | 38.7 | 38.7 | 100.0 |
| | Total | 181 | 100.0 | 100.0 | |

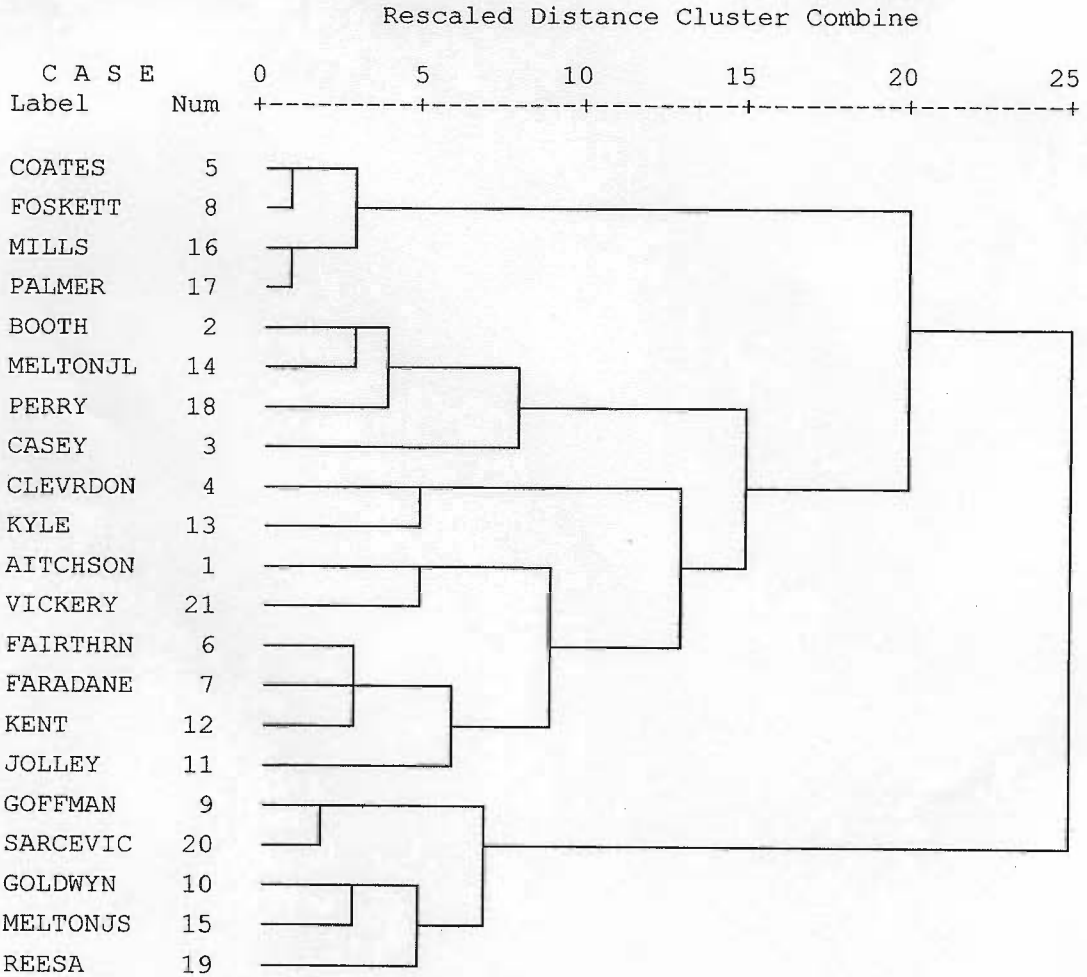
## SPSS 11.0 Select Cluster Analysis output

**Agglomeration Schedule**

| Stage | Cluster Combined | | Coefficients | Stage Cluster First Appears | | Next Stage |
|---|---|---|---|---|---|---|
| | Cluster 1 | Cluster 2 | | Cluster 1 | Cluster 2 | |
| 1 | 5 | 8 | .911 | 0 | 0 | 7 |
| 2 | 16 | 17 | .882 | 0 | 0 | 7 |
| 3 | 9 | 20 | .850 | 0 | 0 | 14 |
| 4 | 6 | 7 | .801 | 0 | 0 | 8 |
| 5 | 2 | 14 | .800 | 0 | 0 | 9 |
| 6 | 10 | 15 | .794 | 0 | 0 | 10 |
| 7 | 5 | 16 | .794 | 1 | 2 | 19 |
| 8 | 6 | 12 | .769 | 4 | 0 | 13 |
| 9 | 2 | 18 | .726 | 5 | 0 | 15 |
| 10 | 10 | 19 | .701 | 6 | 0 | 14 |
| 11 | 4 | 13 | .667 | 0 | 0 | 17 |
| 12 | 1 | 21 | .667 | 0 | 0 | 16 |
| 13 | 6 | 11 | .661 | 8 | 0 | 16 |
| 14 | 9 | 10 | .566 | 3 | 10 | 20 |
| 15 | 2 | 3 | .553 | 9 | 0 | 18 |
| 16 | 1 | 6 | .501 | 12 | 13 | 17 |
| 17 | 1 | 4 | .289 | 16 | 11 | 18 |
| 18 | 1 | 2 | .198 | 17 | 15 | 19 |
| 19 | 1 | 5 | -.042 | 18 | 7 | 20 |
| 20 | 1 | 9 | -.329 | 19 | 14 | 0 |

* * * * * * H I E R A R C H I C A L  C L U S T E R  A N A L Y S I S *
* * * * *

Dendrogram using Complete Linkage



**SPSS 11.0 ALSCAL output**

## Alscal

Iteration history for the 2 dimensional solution (in squared distances)

Young's S-stress formula 1 is used.

| Iteration | S-stress | Improvement |
|---|---|---|
| 1 | .18146 | |
| 2 | .15340 | .02807 |
| 3 | .15080 | .00260 |
| 4 | .15035 | .00045 |

Iterations stopped because
S-stress improvement is less than .001000

Stress and squared correlation (RSQ) in distances

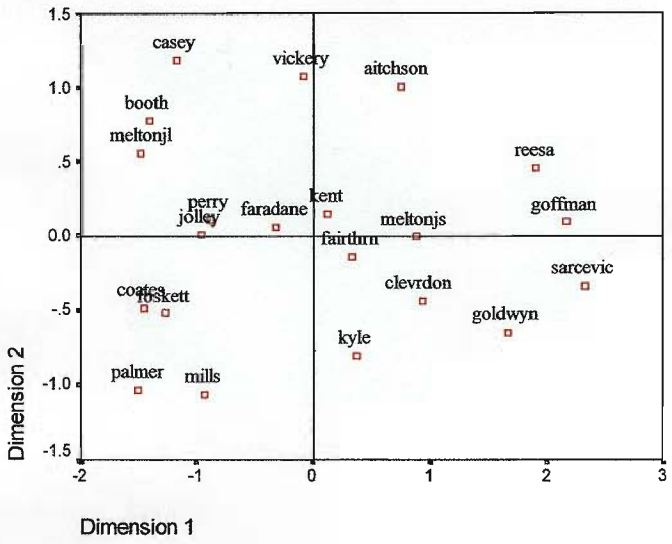
RSQ values are the proportion of variance of the scaled data (disparities)
in the partition (row, matrix, or entire data) which
is accounted for by their corresponding distances.
Stress values are Kruskal's stress formula 1.

For matrix
Stress = .13099 RSQ = .91565

Configuration derived in 2 dimensions

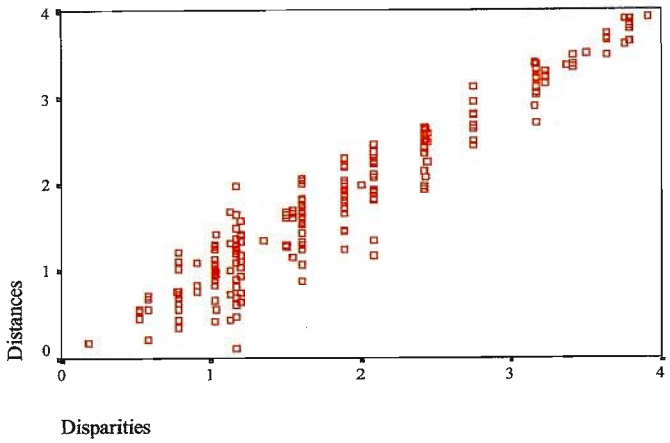
Stimulus Coordinates

| Stimulus Number | Stimulus Name | Dimension 1 | Dimension 2 |
|---|---|---|---|
| 1 | AITCHSON | .7414 | 1.0074 |
| 2 | BOOTH | -1.4069 | .7744 |
| 3 | CASEY | -1.1649 | 1.1812 |
| 4 | CLEVRDON | .9246 | -.4381 |
| 5 | COATES | -1.4481 | -.4825 |
| 6 | FAIRTHRN | .3352 | -.1405 |
| 7 | FARADANE | -.3184 | .0579 |
| 8 | FOSKETT | -1.2694 | -.5203 |
| 9 | GOFFMAN | 2.1639 | .0952 |
| 10 | GOLDWYN | 1.6750 | -.6447 |
| 11 | JOLLEY | -.9561 | .0123 |
| 12 | KENT | .1273 | .1467 |
| 13 | KYLE | .3693 | -.8054 |
| 14 | MELTONJL | -1.4756 | .5598 |
| 15 | MELTONJS | .8787 | -.0042 |
| 16 | MILLS | -.9375 | -1.0662 |
| 17 | PALMER | -1.5052 | -1.0290 |
| 18 | PERRY | -.8834 | .1068 |
| 19 | REESA | 1.9123 | .4609 |
| 20 | SARCEVIC | 2.3294 | -.3409 |
| 21 | VICKERY | -.0917 | 1.0694 |



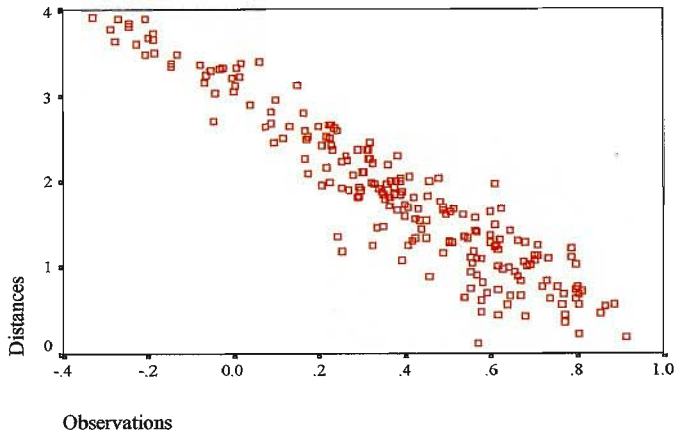
Scatterplot of Linear Fit

Euclidean distance model



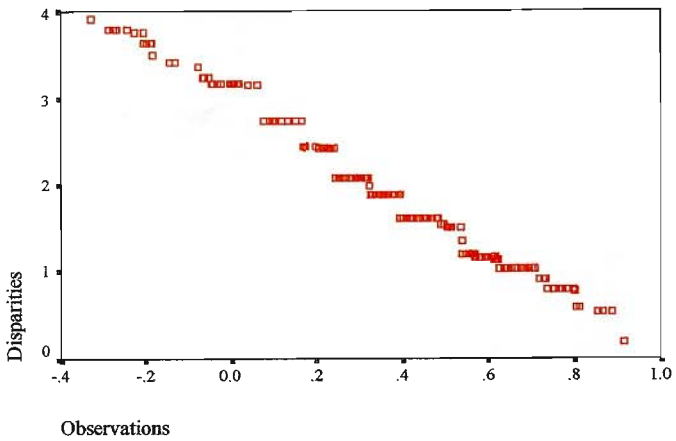

Scatterplot of Nonlinear Fit

Euclidean distance model



Transformation Scatterplot

Euclidean distance model

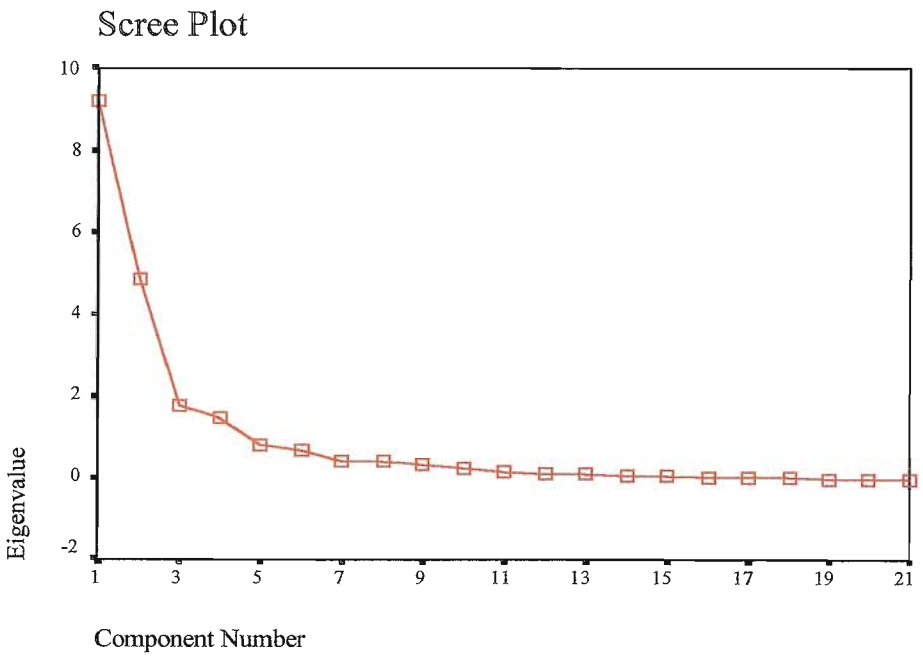


# Factor Analysis

**Communalities**

| | Initial | Extraction |
|---|---|---|
| AITCHSON | 1.000 | .905 |
| BOOTH | 1.000 | .767 |
| CASEY | 1.000 | .733 |
| CLEVRDON | 1.000 | .755 |
| COATES | 1.000 | .885 |
| FAIRTHRN | 1.000 | .852 |
| FARADANE | 1.000 | .893 |
| FOSKETT | 1.000 | .866 |
| GOFFMAN | 1.000 | .850 |
| GOLDWYN | 1.000 | .960 |
| JOLLEY | 1.000 | .821 |
| KENT | 1.000 | .861 |
| KYLE | 1.000 | .638 |
| MELTONJL | 1.000 | .915 |
| MELTONJS | 1.000 | .852 |
| MILLS | 1.000 | .918 |
| PALMER | 1.000 | .919 |
| PERRY | 1.000 | .784 |
| REESA | 1.000 | .652 |
| SARCEVIC | 1.000 | .859 |
| VICKERY | 1.000 | .711 |

Extraction Method: Principal Component Analysis.


Scree Plot
Eigenvalue
10
8
6
4
2
0
-2
1
3
5
7
9
11
13
15
17
19
21
Component Number

**Component Matrix[a]**

| | Component | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| FARADANE | .931 | | | |
| KENT | .898 | | | |
| FAIRTHRN | .876 | | | |
| PERRY | .774 | | | |
| JOLLEY | .759 | | | -.426 |
| MELTONJS | .741 | .536 | | |
| FOSKETT | .722 | -.470 | | |
| KYLE | .691 | | | |
| COATES | .690 | -.536 | | |
| MILLS | .683 | | .528 | |
| AITCHSON | .661 | | | -.493 |
| CLEVRDON | .650 | .437 | | |
| MELTONJL | .645 | -.502 | -.459 | |
| VICKERY | .640 | | | -.507 |
| PALMER | .569 | -.550 | .537 | |
| BOOTH | .561 | | -.431 | |
| CASEY | .529 | | -.525 | |
| SARCEVIC | | .884 | | |
| GOFFMAN | | .844 | | |
| GOLDWYN | .461 | .729 | | .430 |
| REESA | | .714 | | |

Extraction Method: Principal Component Analysis.

a. 4 components extracted.

**Pattern Matrix[a]**

| | Component | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| PALMER | .996 | | | |
| MILLS | .975 | | | |
| COATES | .862 | | | |
| FOSKETT | .836 | | | |
| FARADANE | .447 | | | -.438 |
| GOLDWYN | | 1.039 | | |
| SARCEVIC | | .858 | | |
| CLEVRDON | | .833 | | |
| MELTONJS | | .768 | | |
| GOFFMAN | | .674 | | |
| KYLE | | .657 | | |
| REESA | | .643 | | |
| MELTONJL | | | -.902 | |
| BOOTH | | | -.896 | |
| CASEY | | | -.892 | |
| PERRY | | | -.694 | |
| AITCHSON | | | | -.891 |
| VICKERY | | | | -.809 |
| JOLLEY | .496 | | | -.647 |
| KENT | | | | -.478 |
| FAIRTHRN | | .468 | | -.471 |

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 11 iterations.

**Structure Matrix**

| | Component | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| PALMER | .948 | | | |
| COATES | .931 | | -.548 | |
| MILLS | .930 | | | |
| FOSKETT | .916 | | -.560 | |
| FARADANE | .703 | .426 | -.629 | -.697 |
| GOLDWYN | | .959 | | |
| MELTONJS | | .867 | | -.565 |
| SARCEVIC | | .840 | | |
| CLEVRDON | | .821 | | |
| GOFFMAN | | .783 | | -.517 |
| REESA | | .733 | | -.469 |
| KYLE | .448 | .684 | | |
| MELTONJL | .485 | | -.946 | |
| BOOTH | | | -.874 | |
| CASEY | | | -.845 | |
| PERRY | .577 | | -.827 | |
| AITCHSON | | .461 | | -.933 |
| VICKERY | | | | -.826 |
| KENT | .473 | .573 | -.599 | -.745 |
| FAIRTHRN | .489 | .673 | -.404 | -.743 |
| JOLLEY | .673 | | -.462 | -.730 |

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

**Component Correlation Matrix**

| Component | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.000 | 7.569E-02 | -.465 | -.237 |
| 2 | 7.569E-02 | 1.000 | -7.17E-02 | -.370 |
| 3 | -.465 | -7.17E-02 | 1.000 | .259 |
| 4 | -.237 | -.370 | .259 | 1.000 |

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

## REFERENCES


Abraham, K. (1962). Computer's library role is defended. *The Plain Dealer*.

Afolabi, M. (1983). The Literature of bibliographic classification: a citation study to determine the core literature. (UMI no 841557).

Aldenderfer, M.S. & Blashfield, R.K. (1984). Cluster analysis. (Quantitative Approaches in the Social Sciences No. 44) Beverly Hills: Sage Publications.

Atwood, W.W. (1959). Preface. In Proceedings of the International Conference on Scientific Information, Washington, D.C., November 16-21, 1958, sponsored by National Science Foundation, National Academy of Sciences--National Research Council, American Documentation Institute. Washington, D.C. : National Academy of Sciences--National Research Council.

Austin, D. (1969). Prospects for a new general classification. *Journal of Librarianship*, vol. 1, no. 3, pp. 149-169.

Austin, D. (1972). The CRG research into a freely faceted scheme. In Classification in the 1970's: a second look. Arthur Maltby (ed.), pp. 158-194. London : Clive Bingley.

Baeza-Yates, R. & Ribeiro-Neto, B. (1999). Modern information retrieval. New York : ACM Press.

Bayer, A.E., Smart, J.C., & McLaughlin, G.W. (1990). Mapping intellectual structure of a scientific subfield through author co-citations. *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 444-452.

Borgman, C.L. (2000). Scholarly communication and bibliometrics revisited. In B. Cronin & H. B. Atkins (Eds.), *The Web of Knowledge: a festschrift in honor of Eugene Garfield* (pp. 143-162). Medford, N.J. : Information Today, Inc.

Bowles, M.D. (1999). Crisis in the information age? How the information explosion threatened science, democracy, the library and the human body, 1945-1999. (Doctoral Dissertation, Case Western Reserve University, 1999, UMI No. 9922488).

Braam, R.R., Moed, H.F., van Raan, A.F.J. (1991a). Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for Information Science*, vol. 42, no. 4, pp. 233-251.

Braam, R.R., Moed, H.F., van Raan, A.F.J. (1991b). Mapping of science by combined

co-citation and word analysis. II. Dynamical aspects. *Journal of the American Society for Information Science*, vol. 42, no. 4, pp. 252-266.

Brewer, J. & Hunter, A. (1989). Multimethod research : a synthesis of style. Newbury Park : Sage Publications.

Briton devalues computer. (1962, April 19). *The Plain Dealer*, p. 19.

Brookes, B.C. (1974). Robert Fairthorne and the scope of information science. *Journal of Documentation*, vol. 30, no. 2, pp. 139-152.

Brookes, B.C. (1986). Farradane and relational indexing. *Journal of Information Science Principles and Practice*, vol. 12, nos. 1 & 2, pp. 15-21.

Brownson, H. (2001). Organizing information after Sputnik. Retrieved from the *Pioneers for Information Science* [website] at http://www.libsci.sc.edu/bob/isp/scrapbook.htm.

Bryant, F.B. & Yarnold, P.R. (1997). Principal components analysis and exploratory and confirmatory factor analysis. In L.G. Grim & P.R. Yarnold (Eds.), *Reading and Understanding Multivariate Statistics* (pp. 99-136). Washington, D.C.: American Psychological Association.

Buckland, M. & Hahn, T.B. (1998). Introduction. Historical Studies in Information Science. Medford, N.J. : Information Today, Inc.

Burnett, G., Besant, M. & Chatman, E.A. (2001). Small worlds: normative behavior in virtual communities and feminist bookselling. *Journal of the American Society for Information Science and Technology*, vol. 52, no. 7, pp. 536-547.

Bush, V. (1945). As we may think. *Atlantic Monthly*, vol. 176, no. 1, pp. 101-108.

A Center for Documentation and Communication Research at Western Reserve University. (1955). *American Documentation*, vol. 6, no. 3, pp. 178.

Chatman, E.A. (1986). Diffusion theory: a review and test of a conceptual model in information diffusion. *Journal of the American Society for Information Science*, vol. 37, no. 6, pp. 377-386.

Chatman, E.A. (1990). Alienation theory: application of a conceptual framework to a study of information among janitors. *Research Quarterly*, vol. 29, no. 2, pp. 355-368.

Chatman, E.A. (1991). Life in a small world: applicability of gratification theory to information-seeking behavior. *Journal of the American Society for Information Science*, vol. 42, no. 6, pp. 439-449.

Chatman, E.A. (1992). The Information world of retired women. Westport, Conn.: Greenwood Press.

Chatman, E.A. (1996). The Impoverished life-world of outsiders. *Journal of the American Society for Information Science*, vol. 47, no. 3, pp. 193-206.

Chatman, E.A. (1999). A Theory of life in the round. *Journal of the American Society for Information Science*, vol. 50, no. 3, pp. 207-217.

Classification Research Group (1955). The need for a faceted classification as the basis for all methods of information retrieval. *Library Association Record*, vol. 57, no. 7., pp 262-268.

Classification Research Group (1956a). Bibliography of papers on classification and allied subjects. *Journal of Documentation*, vol. 12, p. 227-230.

Classification Research Group (1956b). Classification Research Group Bulletin No. 1. Unpublished bulletin, Library Association, London.

Classification Research Group (1958). Classification Research Group Bulletin No. 4. *Journal of Documentation*, vol. 14, no. 3, p. 136-143.

Classification Research Group (1959). Classification Research Group Bulletin No. 5. *Journal of Documentation*, vol. 15, no. 1, p. 39-57.

Classification Research Group (1961). Classification Research Group Bulletin No. 6. *Journal of Documentation*, vol. 17, no. 3, p. 156-172.

Classification Research Group (1962). Classification Research Group Bulletin No. 7. *Journal of Documentation*, vol. 18, no. 2, p. 65-88.

Classification Research Group (1964). Classification Research Group Bulletin No. 8. *Journal of Documentation*, vol. 20, no. 3, p. 146-169.

Classification Research Group (1968). Classification Research Group Bulletin No. 9. *Journal of Documentation*, vol. 24, no. 4, p. 273-298.

Classification Research Group (1969). Classification and Information Control. Papers representing the work of the Classification Research Group during 1960-1968. London: Library Association.

Cleverdon, C. (1997). The time of Dorking. In Alan Gilchrist (Ed.), *From Classification to "Knowledge Organization": Dorking Revisited or "Past is Prelude"* (p vi). The Hague Netherlands : FID.

Coates, E.J. (1997). 1957-1997 *Plus CA Change…*what has happened to classification since the Dorking Conference? In Alan Gilchrist (Ed.), *From Classification to "Knowledge Organization": Dorking Revisited or "Past is Prelude"* (p vii). The Hague Netherlands : FID.

Coblans, H. (1974). R.A.F.: a brief biographical note and an annotated bibliography. *Journal of Documentation*, vol. 30, no. 2, pp. 129-138..

Conference on the practical utilization of recorded knowledge at Western Reserve University. (1955). *American Documentation*, vol. 6, no. 4, pp. 255-258.

Curtis-Wright, H. (1988). Jesse Shera, librarianship, and information science. School of Library and Information Services (Occasional research paper, no. 5). Brigham Young University.

Dahlberg, I. (1982). International classification and indexing bibliography: classification systems and thesauri, 1950-1982 (ICIB 1, FID-Publ. 610). Frankfurt, Germany : Indeks Verlag.

Dahlberg, I. (1995). The Future of classification in libraries and networks, a theoretical point of view. *Cataloging and Classification Quarterly*, v. 21, n. 2, p. 23-35.

East, H. (1993). Brian Vickery. In World Encyclopedia of Library and Information Services, 3rd. ed., pp. 855-856. Chicago : American Library Association.

Everitt, B.S., Landau, S. & Leese, M. (2001). Cluster analysis. 4th ed. London: Arnold.

Farkas-Conn, I.S. (1990). From documentation to information science: the beginnings and early development of the American Documentation Institute-American Society for Information Science. New York: Greenwood Press.

Fairthorne, R.A. (1962). An Outsider inside information: USA 1961-1962. *Aslib Proceedings*, vol. 14, no. 11, pp. 280-291.

Farradane, J.E.L. (1952). A Scientific theory of classification and indexing: further considerations. *Journal of Documentation*, vol. 8, no. 2, pp. 83-99.

Farradane, J.E.L. (1957). Classification and mechanical selection. In Proceedings of the International Study Conference on Classification for Information Retrieval, held at Beatrice Webb House, Dorking, England, 13th-17th May 1957. London: Aslib.

Fazel, S. & Danesh, J. (1995). Baha'i scholarship: an examination using citation Analysis [Electronic version]. *Baha'i Studies Review*, vol. 5, no. 1. Retrieved June 15, 2002 from, http://bahai-library.org/articles/citation.analysis.html

Feibleman, J.K. (1985). Theory of integrative levels. In Chan, L.M., Richmond, P.A., & Svenonius, E. (Eds.), Theory of subject analysis: a sourcebook. Littleton, CO. : Libraries Unlimited.

Foskett, A. C. (1973). The Universal Decimal Classification: the history, present status, and future prospects of a large general classification scheme. Hamden, Conn.: Linnet Books.

Foskett, D.J. (1961). Classification and integration levels. In Sayers Memorial Volume: essays in librarianship in memory of William Charles Berwick Sayers. D.J. Foskett and B.I. Palmer (eds.). London : Library Association, pp 136-150.

Foskett, D.J. (1962). The Classification Research Group 1952-1962. *Libri*, v. 12, no. 2, p. 127-138.

Foskett, D.J. (1963). Origins of the conference. In Some problems of a general classification scheme: report on a conference held in London, June 1963 (pp. 11-16). London: Library Association.

Foskett, D. J. (1964). Science, humanism, and libraries. New York : Hafner Publishing.

Foskett, D.J. (1969). Origins of the conference. In Classification and information control: papers representing the work of the Classification Research Group during 1960-1968. Library Association Research Publications, no. 1. London : Library Association.

Foskett, D.J. (1971). The Classification Research Group 1952-1968. In Encyclopedia of Library and Information Science, Vol. 5. Kent, A. & Lancour, H. (Eds.). New York : Marcel Dekker.

Foskett, D.J. (1978). The theory of integrative levels and its relevance to the design of information systems. *Aslib Proceedings*, vol. 30, no. 6, pp. 202-208.

Foskett, D.J. (1988). Brian Vickery: a personal memoir. *Journal of Documentation*, vol. 44, no. 3, pp. 199-204.

Frohmann, B. P. (1983). An investigation of the semantic bases of some theoretical principles of classification proposed by Austin and the CRG. *Cataloging and Classification Quarterly*, vol. 4, no. 11, pp, 11-27.

Furner, J. (2000, July). The Classification Research Group, 1952-2000: a citation analysis. Paper delivered at the Sixth International ISKO Conference, Toronto.

Furner, J. (2001). A Citation study of the work of the CRG. Paper delivered at the ASIS&T Annual Meeting, Washington, D.C., November 6, 2001.

Garfield, E. (1979). Citation indexing: its theory and application in science, technology, and humanities. Philadelphia: ISI Press.

Gilchrist, A. (1997). Introduction. In Alan Gilchrist (Ed.), *From Classification to "Knowledge Organization": Dorking Revisited or "Past is Prelude"* (p vii). The Hague Netherlands : FID.

Goffman, W. & Newill, V. (1964). Generalization of epidemic theory: an application to the transmission of ideas. *Nature,* vol. 204, pp. 225-228.

Griffith, B. C. (1990). Understanding science: studies of communication and information. In Scholarly Communication and Bibliometrics, pp. 31-45. C. L. Borgman (ed.). Newbury Park : Sage Publications.

Griffith, B.C. & Small, H.G., et.al. (1974). The Structure of scientific literatures II: toward a macro- and microstructure of science. *Science Studies,* vol. 4, pp. 339-365.

Harrison, K.C. (1993). Foskett, D. J. In World Encyclopedia of Library and Information Services. 3rd. ed, pp. 299-300. Chicago : American Library Association.

Haslam, D.D. (1975). Library Association. In Encyclopedia of Library and Information Science. (Vol. 14, p. 312-337.) Kent, A. & Lancour, H. (Eds.). New York : Marcel Dekker.

HERE (1971, July 30). [Documentation centre absorbed]. Case Western Reserve University Archives, Press Releases-Publicity, Box 1 45A.

Hernon, P. (1994). Statistics: a component of the research process. Norwood, N.J. : Ablex Publishing Company.

Hewitt, J.A. (1993). F. Wilfred Lancaster. In World Encyclopedia of Library and Information Services. 3rd. ed, p 434. Chicago : American Library Association.

Hjorland, B. (1999). [Review of the book *The DDC, the Universe of Knowledge, and the Post-Modern Library*]. *Journal of the American Society for Information Science*, vol. 50, no. 5, pp. 47-478.

Hopkins, F. (1973). "General Classification Theory: A review of Classification Research Group Work. *Library Resources & Technical Services*, v. 17, n.2., p. 201-210.

Huckaby, S.A.S. (1972). An Enquiry into the theory of integrative levels as the basis for a generalized classification scheme. *Journal of Documentation,* vol. 28, no. 2, pp. 97-106.

Hunter, E.J. (1987). "The Untied Kingdom contribution to subject cataloguing and classification since 1945". *International Cataloguing*, vol. 16, no. 3.

International Society for Knowledge Organization. (2001). [official website]. Retrieved from http://is.gseis.ucla.edu/orgs/isko/isko.html

Justice, A. (2001). "Classification Research Group (CRG) Origins". Unpublished.

Keen, M. (1998). "Cyril W. Cleverdon". *Journal of Documentation*, vol. 54, no. 3, pp. 265-273.

Kent, A. (Ed.). (1960-1961). Information Retrieval and Machine Translation. *Advances in Documentation and Library Science, Vol. 3, pts. 1-2*. Based on the International Conference for Standards on a Common Language for Machine Searching and Translation, Western Reserve University, Cleveland, Ohio, September 6-12, 1959. New York : Interscience Publishers, Inc.

Kent, A. (1961). "Documentation and Communication Research". *Wilson Library Bulletin*, v. 35, no. 10, pp. 772-781.

Kent, A. & Perry, J. (1958). Tools for machine literature searching. New York : Interscience Publishers.

Kline, P. (1994). An Easy Guide to Factor Analysis. London: Routledge.

Koll, M. (1999). *Track Overview*. From 1999 American Society for Information Science Annual Conference. (Cassette recording SIS 969). Washington, D.C. : ASIS.

Kreuzman, H. B. (1990). Theories of rationality in epistemology and philosophy of science. (UMI no. 9033096)

Kuhn, T. (1970). The Structure of Scientific Revolutions. 2nd ed. Chicago : University of Chicago Press.

Kuhn, T. (1977). "The Essential Tension: tradition and innovation in scientific research". In The Essential Tension: selected studies in scientific tradition and change (pp. 225-239). Chicago : University of Chicago Press.

Kyle, B.R.F. (1956). Privilege and public provision in the intellectual welfare state. *Journal of Documentation*, vol. 12, no. 1, pp. 15-23.

Kyle, B.R.F. (1963). Some searching questions on documentation research. *American Documentation*, vol. 4, no. 2., p. 93.

Lancaster, F.W. (1993). Cyril Cleverdon. In World Encyclopedia of Library and

Information Services. 3rd. ed., pp. 212-213. Chicago : American Library Association.

Lievrouw, L.A. (1990). "Reconciling structure and process in the study of scholarly communication". In Scholarly Communication and Bibliometrics, pp. 59-69. C. L. Borgman (ed.). Newbury Park : Sage Publications.

Lloyd, G. A. (1976). Universal Decimal Classification. In Classification in the 1970's: a second look. Arthur Maltby (Ed.). London: Clive Bingley, pp. 99-118.

Locke, W.N. (1975). Machine translation. In *Encyclopedia of Library and Information Science*. (Vol. 16). New York : Marcel Dekker.

McCain, K.W. (1983). "The author co-citation structure of macroeconomics". *Scientometrics*, vol. 5, pp. 277-289.

McCain, K.W. (1984). "Longitudinal author co-citation mapping: the changing structure of Macroeconomics". *Journal of the American Society for Information Science*, vol. 35, no. 6, pp. 351-359.

McCain, K.W. (1986a). "The paper trails of scholarship: mapping the literature of genetics". *Library Quarterly*, vol. 56, no. 3, pp. 258-271.

McCain, K.W. (1986b). "Cocited author mapping as a valid representation of intellectual structure". *Journal of the American Society for Information Science*, vol. 37, no. 3, pp. 111-122.

McCain, K.W. (1988). "Evaluating cocited author search performance in a collaborative specialty". *Journal of the American Society for Information Science*, vol. 39, no. 6, pp. 428-431.

McCain, K.W. (1989). "Mapping authors in intellectual space: population genetics in the 1980s. *Communication Research*, vol. 16, pp. 667-681.

McCain, K.W. (1990). "Mapping authors in intellectual space: a technical overview". *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 433-443.

McCain, K.W. (1991). "Mapping economics through journal literature: an experiment in journal cocitation analysis". *Journal of the American Society for Information Science*, vol. 42, no. 4, pp. 290-296.

McIlwaine, I. & Broughton, V. (2000). "The Classification Research Group-- then and now". *Knowledge Organization*, vol. 27, no. 4, pp. 195-199.

Markey, K. & Demeyer, A.N. (1986). Final report to the Council on Library Resources: Dewey Decimal Classification online project: evaluation of a library schedule and index integrated into the subject searching capabilities of an online catalog. (Report Number: OCLC/OPR/RR-86/1). Dublin, Ohio : OCLC.

Melton, J.S. (1960). A Note on the compatibility of two information systems: colon classification and Western Reserve University, etc. (Technical Report no. 13). Cleveland: Western Reserve University, Center for Documentation and Communication Research.

Miksa, F. (1984). The Development of classification at the Library of Congress. University of Illinois, Graduate School of Library and Information Science Occasional Papers, no. 164.

Miksa, F. (1987, March 30). Research patterns and research libraries. Text of an address given at the Sixth Annual Conference of Directors of Research Libraries in OCLC, Dublin, OH.

Miksa, F. (1998). The DDC, the universe of knowledge, and the post-modern library. Albany, N.Y.: Forest Press.

Miles, M.B. & Huberman, A.M. (1994). Qualitative data analysis : an expanded sourcebook. 2nd ed. Thousands Oaks : Sage Publications.

Mills, J. (1997). Comments on Dorking and after. In Alan Gilchrist (Ed.), *From Classification to "Knowledge Organization": Dorking Revisited or "Past is Prelude"* (p xi). The Hague Netherlands : FID.

Mills, J. (1998). Classification for retrieval. Retrieved from the *Pioneers for Information Science* [website] at http://www.libsci.sc.edu/bob/isp/scrapbook.htm

Mitchell, J. (1995). "DDC 21 and beyond; the Dewey Decimal Classification prepares for the future." *Cataloging and Classification Quarterly*, v. 21, n. 2, p.37-45.

Morville, P. & Rosenfeld L. (1998). Information architecture for the world wide web. California: O'Reilly & Associates.

Paisley, W. (1990). "An Oasis where trails cross: the improbably cocitation networks of a multidiscipline." *Journal of the American Society for Information Science*, v. 41, n. 6, pp. 459-468.

The Pearson Product Moment Correlation Coefficient (Pearson's r) plus various & sundry issues. (n.d.). Retrieved May 13, 2002 from http://psychology.wadsworth.com/workshops/correlation.html

Perry, J. & Kent, A. (1957). Documentation and Information Retrieval: an introduction

to basic principles and cost analysis. Cleveland : Western Reserve University Press.

Perry, J. & Kent, A. (1958). Tools for machine literature searching. *Library Science and Documentation, Vol. 1.* New York : Interscience Publishers, Ltd.

Price, D. (1965). "Networks of scientific papers". *Science*, vol. 149, July 1965, pp. 510-15.

Proceedings of the International Conference on Scientific Information, Washington, D.C, November 16-21, 1958. (1959). Washington, D.C. : National Academy of Sciences, National Research Council.

Proceedings of the International Study Conference on Classification for Information Retrieval held at Beatrice Webb House, Dorking, England, 13th-17th May 1957. (1957). London: Aslib.

Puri, K.I. & Arora, R.L. (1988). Professor Allen Kent's contribution in the field of librarianship and documentation. *Herald of Library Science,* vol. 27, nos. 1-2, pp. 24-27.

Rees, A.M. (1963). Semantic factors, role indicators, et alia: eight years of information retrieval at Western Reserve University. *Aslib Proceedings,* vol. 15, no. 12, pp. 350-363.

Richmond, P. (1988). "Precedent-setting contributions to modern classification." *Journal of Documentation,* vol. 44, no. 3, pp. 242-249.

Rosenfeld, L. & Morville, P. (1998). Information Architecture for the World Wide Web. Sebastopol, CA. : O'Reilly & Associates.

Rowley, J. (1992). Organizing knowledge: an introduction to information retrieval. Vermont : Gower.

Salton, G. & McGill, M.J. (1983). Introduction to modern information retrieval. New York : McGraw-Hill Book Company.

Sandstrom, P.E. (1998). Information foraging among anthropologists in the invisible college of human behavioral ecology: an author co-citation analysis. (Doctoral Dissertation, Indiana University, UMI no. 9919485).

Saracevic, T. (1964). The Comparative Systems Laboratory at Western Reserve University: progress report on operational procedures. In Elias, A.W. (Ed.), *Proceedings of the American Documentation Institute: Vol. 1. Parameters of Information Science Annual Meeting, Philadelphia, Pennsylvania, October 5-8, 1964,* (pp. 427-431). American Documentation Institute.

Shera, J. (1971). "Case Western Reserve University, School of Library Science". In Encyclopedia of Library and Information Science, Vol. 4. Kent, A. & Lancour, H. (Eds.). New York : Marcel Dekker.

Shera, J., Kent, A., & Perry, J. (Eds.). (1956). Documentation in Action. Based on 1956 Conference on Documentation at Western Reserve University. New York : Reinhold Publishing Corp.

Shera, J., Kent, A., & Perry, J. (Eds.). (1957). Information Systems in Documentation. *Advances in Documentation and Library Science*: *Vol. 2*. Based on the Symposium on Systems for Information Retrieval held at Western Reserve University, Cleveland, Ohio, in April 1957. New York : Interscience Publishers, Ltd.

Small, H.G. (1976). "Structural Dynamics of Scientific Literature". *International Classification*, vol. 3, no. 2, pp. 67-74.

Small, H.G. (1977). "A Co-citation model of a scientific specialty : a longitudinal study of collagen research." *Social Studies of Science*, vol. 7, pp. 139-166.

Small, H.G. (1980). "Co-citation context analysis and the structure of paradigms." *Journal of Documentation*, vol. 36, no. 3, pp. 183-196.

Small, H.G. (1981). "The Relationship of information science to the social sciences: a co-citation analysis." *Information Processing and Management*, vol. 17, pp. 39-50.

Small, H.G. (1986). "The Synthesis of specialty narratives from co-citation clusters." *Journal of the American Society for Information Science*, vol. 37, no. 3, pp. 97-110.

Small, H.G. & Crane, D. (1979). "Specialties and disciplines in science and social science: an examination of their structure using citation indexes." *Scientometrics*, vol. 1, no. 5-6, pp. 445-461.

Small, H.G. & Griffith, B.C. (1974). "The Structure of scientific literatures I: identifying and graphing specialties". *Science Studies*, vol. 4, pp. 17-40.

Spiteri, L. (1995). "The Classification Research Group and the theory of integrative levels". [Electronic version]. Retrieved online at http://www.lis.uiuc.edu/review/summer1995/spiteri.html

Soergel, D. (2001). "ASIST SIG/CR Classification Workshop 2000: classification for user support and learning". Bulletin for the American Society for Information Science and Technology, vol. 27, no. 4, April/May 2001, pp. 15-21.

Snow, C.P. (1959). Two cultures and a second look. Cambridge: Cambridge University Press.

Stalans, L.J. (1995). Multidimensional scaling. In L.G. Grim & P.R. Yarnold (Eds.), *Reading and Understanding Multivariate Statistics* (pp. 137-168). Washington, D.C.: American Psychological Association.

Van Raan, A.F.J. & van Leeuwen, Th. N. (2001). Assessment of the scientific basis of interdisciplinary, applied research: application of bibliometric methods in nutrition and food research [Electronic version]. Retrieved August 29, 2001, from http://sahara.fsw.leidenuniv.nl/cwts/tnopubl.html

Vickery, B.C. (1953). Classification Research Group. *Library Association Record*, vol. 55, no. 6, pp. 187-188.

Vickery, B.C. (1959). The Structure of semantic coding, a review. *American Documentation*, vol. 10, pp. 234-241.

Vickery, B.C. (1960). Faceted classification: a guide to construction and use of special schemes. London : ASLIB.

Vickery, B.C. (1966). Faceted classification schemes. In Artandi, S. (Ed.) Rutgers Series on Systems for the Intellectual Organization of Information (Vol. 5). New Brunswick, N.J. : Graduate School of Library Service, Rutgers, the State University.

Walker, P.B. (1974). Robert Arthur Fairthorne: an appreciation. *Journal of Documentation*, vol. 30, no. 2, pp. 127-128.

Weber, R.P. (1990). Basic content analysis. 2nd ed. Series: Quantitative Applications in the Social Sciences. Newbury Park, CA.: Sage Publications.

White, H.D. & Griffith (1981). Author co-citation: a literature measure of intellectual structure. *Journal of the American Society of Information Science*, vol. 32, pp. 163-171.

White, H.D.& Griffith B.C. (1982). Authors as markers of intellectual space: co-citation in studies of science, technology, and society. *Journal of Documentation*. Vol. 38, no. 4, pp. 255-272.

White, H.D. & McCain, K.W. (1998). Visualizing a discipline: an author co-citation analysis of information science, 1972-1995. *Journal of the American Society of Information Science*, vol. 49, no. 4, pp. 327-355.

Wilson, L. (1965). Barbara Kyle. *Journal of Documentation*, vol. 21, no. 4, pp. 227-235.