

Characterising scholarly identities :a citation identity analysis of the field of the scientific study of consciousness

Author:

Orsatti, Joanne

Publication Date:

2007

DOI:

<https://doi.org/10.26190/unsworks/6584>

License:

<https://creativecommons.org/licenses/by-nc-nd/3.0/au/>

Link to license to see what you are allowed to do with this resource.

Downloaded from <http://hdl.handle.net/1959.4/40472> in <https://unsworks.unsw.edu.au> on 2024-04-25

UNSW



>013751689

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES

Thesis/Dissertation Sheet

Surname or Family name: ORSATTI

First name: Joanne

Abbreviation for degree as given in the University calendar: MInfStuds

School: School of Information Technology and Management (SISTM)

Title: Characterising scholarly identities: a citation identity analysis of the field of the Scientific Study of Consciousness

Other name/s:

Faculty: Business

Abstract 350 words maximum: (PLEASE TYPE)

The professional profile of researchers is established through communication of scientific work practices, leading to the establishment of a scholarly identity. Understanding scholarly identities is currently addressed through a conceptualisation of research narrative mechanisms. Citation and citing practices are a central component of scientific communication work practices. Therefore understanding these formal communication practices of researchers through their citing behaviours may contribute to the building of scholarly identity. This study is undertaken to understand whether scholarly identity could be informed through the use of citation identities. Studies on the citation identities of individuals were conducted, using authors working in the area of Consciousness, which provided a diverse field of participants for the testing of citation analysis techniques. This is accomplished through methodological development and further examined using a combination of field-level and individual-level analyses. A new methodology was developed for the generation of citing identities, based on the calculation of the Gini coefficient and the citee-citation ratio of authors' citing profiles. The resulting relationship was found to have high levels of consistency across a heterogeneous set of researchers. An exploration of identification of author characteristics was subsequently undertaken using the new methodology and existing citation analysis techniques. The techniques were successful in identifying departures from conventional citation practice, highlighting idiosyncrasies well, but otherwise understanding of scholarly identity through citation analysis was only marginally successful. A portion of the difficulty of achieving clarity was the complexity of the Consciousness author set, which was useful for establishing broad applicability of a new methodology, but poor for judging its successful application. In summary, definition of citing identity type offers possibilities for improving the understanding of scholarly identity, but will require further methodology development to reach its full potential.

Declaration relating to disposition of project thesis/dissertation
I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.
I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).

Signature

Witness

Date

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

FOR OFFICE USE ONLY

Date of completion of requirements for Award:

CHARACTERISING SCHOLARLY IDENTITIES: A CITATION IDENTITY ANALYSIS OF THE FIELD OF THE SCIENTIFIC STUDY OF CONSCIOUSNESS

JOANNE ORSATTI

A thesis submitted in fulfillment of the requirements for the degree of
Masters of Information Studies (Research)

SCHOOL OF INFORMATION SYSTEMS, TECHNOLOGY AND
MANAGEMENT
UNIVERSITY OF NEW SOUTH WALES

May 2007

UNSW
17 DEC /
LIBRARY

ACKNOWLEDGEMENTS

I would like to acknowledge and thank Connie Wilson, Mari Davis and the BIRGers, both home grown and visiting, for introducing me to the vast vista of research, and for those very exciting early days. All have provided me with many opportunities to learn both intellectually and socially. Mari has shared with me her passion for studying people and the ways they make knowledge. Connie has provided among many things the intriguing world of Consciousness researchers, much patience and some much appreciated financial support.

Maureen Henninger, cannot be thanked enough for her mentorship, incisive mind, for being an excellent barometer and for chook-in-milk that is always accompanied by excellent conversation. Theresa Anderson for passionately living with relevance and for taking me along for some of the rides. Margaret Lo for foremost being Margaret, but also for her unerring support, help and for always going above and beyond.

I would also like to thank fellow 'brothers-in-arms': Jonathan Paul for always having an interesting life as well as a fine set of enthusiast quality ideas, Paul Scifleet for many surprisingly fruitful conversations about his work which he managed to successfully combine with being a very good idea litmus paper for my own, and to Weiping Yue for showing it could be done.

A deep and heartfelt acknowledgement and thanks to Jochen Gläser for working with me through various stages of and changes in my research. He introduced me to many foundational ideas and challenged / inspired me to continuously do better research.

Most of all I would like to thank my family and friends, I do not have room to list you individually, but value all you have done to keep me going, you all know who you are, so many, many thanks. I particularly want to acknowledge the loving and patient support of my parents, Guido and Gloria, as well as Michelle, Amy, Mike, Greg, Rachel, Ben and Margaret Beehag. Finally and most completely I would like to thank Andrew and Nathan who have truly made the thesis possible and have unfailingly been there throughout, through both the thick and the very thin. I am a very lucky person to know you all, especially my fellas. Many Thanks.

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed .

Date

COPYRIGHT STATEMENT

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only).

I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed

Date

AUTHENTICITY STATEMENT

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed/.....

Date

ABSTRACT

The professional profile of researchers is established through communication of scientific work practices, leading to the establishment of a scholarly identity.

Understanding scholarly identities is currently addressed through a conceptualisation of research narrative mechanisms. Citation and citing practices are a central component of scientific communication work practices. Therefore understanding these formal communication practices of researchers through their citing behaviours may contribute to the building of scholarly identity.

This study is undertaken to understand whether scholarly identity could be informed through the use of citation identities. Studies on the citation identities of individuals were conducted, using authors working in the area of Consciousness, which provided a diverse field of participants for the testing of citation analysis techniques. This is accomplished through methodological development and further examined using a combination of field-level and individual-level analyses.

A new methodology was developed for the generation of citing identities, based on the calculation of the Gini coefficient and the citee-citation ratio of authors' citing profiles. The resulting relationship was found to have high levels of consistency across a heterogenous set of researchers.

An exploration of identification of author characteristics was subsequently undertaken using the new methodology and existing citation analysis techniques. The techniques were successful in identifying departures from conventional citation practice, highlighting idiosyncrasies well, but otherwise understanding of scholarly identity through citation analysis was only marginally successful. A portion of the difficulty of achieving clarity was the complexity of the Consciousness author set, which was useful for establishing broad applicability of a new methodology, but poor for judging its successful application.

In summary, definition of citing identity type offers possibilities for improving the understanding of scholarly identity, but will require further methodology development to reach its full potential.

TABLE OF CONTENTS

Acknowledgements ii

Abstract v

Table of Contents v

Table of Figures..... ix

List of Tables xi

1. Introduction 1

2. Citation theory and the situated communicating individual in knowledge
production.....5

2.1. Citers as communicating authors..... 5

2.2. Citation theories 5

2.3. A typology of citation theories..... 6

2.4. The normative citation theory 9

2.5. The tactical citation theory 14

2.6. Empirical tests of the attribution citation theories 20

2.6.1. Operationalisations of the attribution theories of citation and convergent
findings21

2.7. Theoretical convergence between the attribution theories of citation.....30

2.8. Association theories of citation 33

2.8.1. Symbolic ‘theory’ of citation 34

2.8.2. Self-organising communication systems ‘theory’ of citation..... 39

2.9. The situated knowledge producing individual author 44

3. Communicating identity as a scientific work practice - Sociocognitive
representation of the individual knowledge producer..... 46

3.1. Citation identities and stylistics 47

3.1.1. Citation identities and the normative citation theory – White’s analysis50

3.1.2. Content and the sociocognitive determinants of citation identities..... 53

3.2. scholarly identities and research narratives..... 56

3.3. sociocognitive representation and scientific work practices..... 59

3.4. Knowledge production institutions, Intellectual positioning and
representational space..... 64

3.4.1. Communication channels, publication acceptance and intellectual
positioning 67

3.4.2. Research models and the citing individual researcher.....68

3.4.3. Research trails – the role of Problem choice and research stratgies in
representational outcomes 70

3.4.4. Membership, marginality and migration 73

3.5.	Scope of thesis	76
4.	Citee-citation concentration and citing identity types	78
4.1.	Introduction	78
4.2.	The scientific study of consciousness	78
4.2.1.	Introduction to the scientific study of consciousness.....	78
4.2.2.	The scientific study of consciousness as boundary specialty	80
4.2.3.	Establishing researcher participation – defining the data set.....	81
4.2.4.	Collection of biographical author information.....	83
4.2.5.	Collecting citing identities	84
4.3.	Gini as a measure of citee-citation concentration.....	85
4.4.	The Gini versus citee-citation relation	86
4.4.1.	Assumptions and statistical discriptives for the Gini versus type-token relation – appropriateness for linear regression analysis	90
4.4.2.	Mathematical examination of the Gini versus citee-citation relation ..	92
4.4.3.	Examination of Gini versus citee-citation relation and variables through data variation	94
4.4.4.	Implications of Gini versus citee-citation ratio findings	98
4.5.	Gini versus citee-citation relation and the informetric “laws”	99
4.5.1.	Zipf’s law	99
4.5.2.	Power law and logarithmic citation distributions	100
4.6.	Application of the Gini versus citee-citation ratio relation	104
4.7.	Establishing citing identity behaviour types	106
4.8.	Citing identity behaviour types	109
4.9.	Conclusion	110
5.	Citing identity type analysis, research personas and representational space. 112	
5.1.	Introduction	112
5.2.	Research personas and boundary specialty representational space	112
5.3.	Methodology	114
5.3.1.	Collection of bibliographic author data	114
5.3.2.	Research personas of participants in representational space: consciousness boundary specialty and citing identity types	114
5.3.3.	Citing identity type analyses	120
5.4.	Results - Research personas in representational space and citing identity types.....	120
5.4.1.	Intellectual placement of participants by others - author co-citation analysis (ACA).....	120

5.4.2.	Collaborative similarity between authors based on co-author and shared co-author matrix analysis	126
5.4.3.	Communication channel similarity based on journal co-publication matrix analysis	139
5.4.4.	Communication channel similarity and citing identity type.....	146
5.4.5.	Self-placement of participants in representational space - citee matrix analyses compared to citing identity types.....	148
5.5.	Results - Citing identity type and citing behaviour analyses	149
5.5.1.	Productivity.....	149
5.5.2.	Length of publication career.....	150
5.5.3.	Unicitation (unicites)/Recitation (recites)	152
5.5.4.	Self-citing.....	155
5.5.5.	Individual research field affiliation	159
5.6.	Discussion.....	160
5.6.1.	Application of methodology.....	160
5.6.2.	Research persona identification	161
5.7.	Conclusion	162
6.	Conclusions and Recommendations for Further Research.....	163
6.1.	Concluding Remarks	163
6.2.	Recommendations for Further Research	164
	Bibliography.....	166
	Appendix A. Author codes	177
	Appendix B. Program code	178
	Appendix B.1. Code for bibliographic extraction program	178
	Appendix B.2. Program for generating a symmetric matrix.....	186
	Appendix C. Citing identity type cluster solution	192
	Appendix C.1. Table for comparison of alternative cluster analysis results, for a 7 cluster solution.....	192
	Appendix C.2. 7 cluster solution using Ward Method of Cluster Analysis	194
	Appendix D. Details from representational space and citing identity analysis (Chapter 5)	197
	Appendix D.1. Authors that co-author but not within consciousness participant set.....	197
	Appendix D.2. Journals that are published across all journal clusters	197
	Appendix D.3. Journals that are published in by 6 journal clusters	197
	Appendix D.4. Publication career year range for individual authors	199
	Appendix E. Author-citation matrix cluster solution	200
	Appendix F. Co-author matrix cluster solution	203
	Appendix G. Co-journal publication matrix cluster solution	205

Appendix H. Co-citee matrix cluster solution.....	208
---	-----

TABLE OF FIGURES

Figure 2.1 Visualisation of Leydesdorff's "Theories of citation" 43

Figure 4.1 Example of an individual author's normalised cumulative citation and citee distribution 85

Figure 4.2 Gini versus Citee-citation ratio.....87

Figure 4.3 Gini versus log citee-log citation relation 89

Figure 4.4 Placement of the theoretical slope (broken line) when all citations from authors are included95

Figure 4.5 Placement of the theoretical slope (broken line) when citations from authors with a frequency below 5 are removed..... 96

Figure 4.6 Schematic of an author's citation distribution with equal citee-citation ratio and varied Gini values.....97

Figure 4.7 Schematic of citation level versus citees for citation distributions identified in Figure 4.6. 98

Figure 4.8 Power law fit R^2 coefficient versus Gini coefficient100

Figure 4.9 Logarithmic fit R^2 coefficient versus Gini coefficient..... 101

Figure 4.10 The Gini versus citee-citation ratio relation with (■) logarithmic fit better than power law fit, and (◆) power law fit better than logarithmic fit.....102

Figure 4.11 THOMAS NATSOULAS' individual citee-citation plot103

Figure 4.12 FRANCIS CRICK'S individual citee-citation plot.....104

Figure 4.13 Ward's and Centroid cluster analysis compared with the Gini and Citee-citation relation108

Figure 5.1 MDS PROXSCAL solution for Author co-citation matrix with author co-citation clusters overlaid 121

Figure 5.2 MDS PROXSCAL solution dimension 1 in relation to co-citation counts from co-citation matrix..... 122

Figure 5.3 MDS PROXSCAL solution dimension 2 and authors year of birth124

Figure 5.4 MDS PROXSCAL solution for Author co-citation matrix with author co-citation clusters overlaid with citing identity types 125

Figure 5.5 Combined core-shared co-author symmetrical matrix MDS PROXSCAL solution overlaid with 8 cluster Average linkage (within groups) cosine solution127

Figure 5.6 Core author direct co-authorships networks. Numbers indicate the corresponding assignment of authors to co-author clusters. 129

Figure 5.7 Direct core co-author networks labelled with year of receipt of highest degree 133

Figure 5.8 MDS PROXSCAL solution overlaid with citing identity types 134

Figure 5.9 Frequency of occurrence in co-author cluster according to citing identity type.....	136
Figure 5.10 Euclidean distance multidimensional scaling solution for symmetrical author-author journal publication matrix labelled by author	140
Figure 5.11 Linear regression of author-author productivity within Consciousness journal set	141
Figure 5.12 MDS Euclidean distance solution labelled with citing identity types and overlaid with journal content cluster Ward’s solution	142
Figure 5.13 Network representation of connections between journal title keyword topic categories, according to unique journal usage by journal clusters.....	144
Figure 5.14 Citing identity types and their frequency of appearance in journal clusters	146
Figure 5.15 MDS of citee similarity matrix according to citing identity type	148
Figure 5.16 Simple productivity related to Gini measure of citing density by citing identity types.....	150
Figure 5.17 Length of publication career as represented in ISI by citing identity type.....	151
Figure 5.18 Simple productivity in relation to career length in years, by citing identity type.....	152
Figure 5.19 Average of unicites according to citing identity types	153
Figure 5.20 Reciting (citee recited/total recites) patterns according to citing identity types	154
Figure 5.21 Self-cites normalised by total cite yield by citing identity type.....	155
Figure 5.22 Authors that do not self-cite most frequently compared with authors that do, according to simple productivity counts and ISI publication career length.....	156
Figure 5.23 Simple productivity in relation to self-citing activity normalised for total cite yield	157
Figure 5.24 Self-citing patterns normalised for cite yield by ISI publication career length in years	158

LIST OF TABLES

Table 4-1 Comparison of citing identity types represented in the relation109

Table 5-1 Table of known Student-Mentor relationships within the Consciousness data
set138

Table 5-2 Number of journals common between the clusters143

Table 5-3 Journal title content clusters for journals unique to journal cluster143

Table 5-4 Research field affiliation of authors as obtained from biographical accounts
.....159

1. INTRODUCTION

Definitive citation theory formulation, despite long term attention and much considered debate, remains elusive. Foundational aspects of its development are still strongly contested by oppositional forces within its satellite fields of citation studies, information sciences, sociology of science and science policy (Leydesdorff, 1989). The inability of these fields to reach a fundamental understanding and consensus of what in fact a citation represents has limiting implications for the application of citation data, thereby influencing the utility of the potentially vast wealth of data that citation analysis can provide these fields. Citation data allows for broad ranging and unobtrusive insights into the practice of science at multiple levels, as presented through an integral medium for the communication of scientific work. Citation analysis also provides a useful platform for the convergent approaches of quantitative and qualitative analysis of scientific activity. This combination allows for examinations of scientific practice that would otherwise be difficult to obtain. However, as an appropriate formulation of citation theory continues to be unattained, its extension into derivative analyses still requires complex justificatory frameworks to be complete. This study considers citation theory from the perspective gained by the development of the conceptual possibilities of scholarly identities and their relationships to citation identities and citation stylistics.

Over a series of publications Howard White has introduced (2000; 2001b), developed (2001a; 2004b) and utilised (2004a; 2004) the concept of 'citation identities' to the literature of citation theory and analysis. Citation identities are defined by White (2000) in this initial formulation as the 'referential choices that position an author intellectually' (p.479). The fundamental assumption behind this definition and its role in understanding an author's work is that 'any author's main motivation in choosing citees is their perceived relevance to an argument or exposition, regardless of whether social ties exist' (p.484). White's theoretical concern is focused on motivational approaches to citation.¹ This aspect itself forms a fruitful point for further analysis using citation identities. As will be developed in this study, White's conceptualisation and empirical analyses of citation stylistics and identities concurrently provides a strong platform for the examination of functional theories of citation.

¹ This motivational explanatory framework is consistently focused on within White's approach to the theoretical questions raised within his work on citation identities (including his recent studies (2004a; 2004b; 2004); these aspects of the theoretical implications of White's work will be further discussed in Section 3.1.

What is also made clear with the development of the concept of citation identities is that it becomes possible to address the question of the connections between social ties and citing behaviour. This potential is indicated by an observation made by White and is accompanied by a statement concerning the limited knowledge that is currently available concerning the issues raised by citation identities.

‘There is reason to think that the employing organization and the invisible college are extremely fertile in producing recurring citees (see White, 2001b). But we do not have hard ethnographic data from many fields and many authors as evidence’ (White, 2000, p.485).

This tantalising observation is further detailed in White’s later paper on author citation profiling,

‘My guess is that, in the directed relationships, citation is affected by a sense of intellectual seniority. ...there is an apparent tendency to cite up or across the seniority chain much more than down’ (White, 2001a, p.625).

These observations and their accompanying preliminary interpretations demonstrate the possibility of utilising the concept of citation identities to allow for the concurrent examination of three crucial elements within citation behaviour, these elements being content, context and scientific practice in relation to citing activity.

From analysis of use and reuse within an individual authors’ citing patterns White (2001b) begins to examine and raise the interesting prospect of an author’s citation style. Introducing a very simple measure of concentration, the citee-citation ratio, and recitation analysis, White identifies patterns of ‘core-scatter’ within authors’ citation oeuvres. Citation identities and their extension into citation stylistics introduces the potential for examining citation behaviour as it exists within scientific work practice. In this framework authors can be examined within their context of practice, allowing simultaneous consideration of the scientific content of their work, its broader socio-cognitive context and analysis across their publication careers. This represents an opportunity to examine citation behaviour in a way not recognised previously within citation studies.

Emerging from the previous work on citation identities and stylistics, the following questions are further investigated in this study:

1. Can individual authors be differentiated according to their citation identities and types?
2. Are citation identities empirically linked to research practices of individuals?

-
-
3. Do citation theories provide an explanation for the empirical phenomenon of citation identities and types?
 4. What does citation theory need to address in order to provide an account of citation identities?

To achieve these aims distributional citation identity concentration and its implications for understanding citation identities themselves is investigated, along with their relevance to citation theory. A sample size is used that allows for statistical significance testing, as citation identities are a relatively recent addition to citation studies. The resultant concentration measures are used as a point of disambiguation to examine citation stylistics in relation to Lievrouw's (1996) conceptualisation of scholarly identity, and their potential and specific roles within citation theory development. In order to meet the criterion of examining researchers from diverse disciplinary backgrounds, the field chosen for analysis is the 'scientific study of consciousness'. This field incorporates the requirements of being a content-defined scientific community as well as being made up of a range of researchers belonging to a diverse set of disciplines. Further, this analysis also considers some of the other less dominant citation theories, not previously connected to these themes, and how they can be related to the new information available through citation identities and the concept of citation stylistics.

This thesis will address these aims through the following sequence of chapters:

SUMMARY OF CHAPTERS:

CHAPTER 2: CITATION THEORY AND THE SITUATED COMMUNICATING INDIVIDUAL IN KNOWLEDGE PRODUCTION

A framework is developed that allows placement of the concept of citation identities within the citation theory literature. The major theories of citation (normative, tactical, and associative), are therefore addressed to provide a context for later development of individual scientific identity.

CHAPTER 3: COMMUNICATING IDENTITY AS A SCIENTIFIC WORK PRACTICE - SOCIOCOGNITIVE REPRESENTATION OF THE INDIVIDUAL KNOWLEDGE PRODUCER

This chapter places the use of citation theory to identify the individual researcher within broader concepts of scholarly identity, to provide a concept of what citation theories will require to inform understanding of scholarly identity. It is argued that further development of citation theory should be grounded in the functions citation plays within scientific work practices. The development of citation identities is reviewed, followed by a review of theories on researcher identity. This is followed by

examination of sociocognitive representation and work practices, to place the reviewed identity theories within theories on use and form of communication. Finally, the review encompasses the frameworks for scholarly communication and research.

CHAPTER 4: CITEE-CITATION CONCENTRATION AND CITING IDENTITY TYPES

This chapter introduces the case field of the Scientific Studies of Consciousness, and utilises this data set for the development of theory associated with citation identity. Distributional aspects of citation identities are examined in terms of their relative concentration, and extension of this methodology undertaken through examination of citation stylistics. Data collection rationales and descriptions of their implementation are discussed in terms of how they are employed to address the questions raised within the study. A typology of citing identities is then developed to identify possible types of citing profiles to allow for further investigation of the features that may determine the citation behaviour represented.

CHAPTER 5: CITING IDENTITY TYPE ANALYSIS, RESEARCH PERSONAS AND REPRESENTATIONAL SPACE

This chapter outlines the methodologies used to examine citing profiles, including co-citation and citee network analysis. The identification of a number of prototypical individuals is discussed, providing the cases for a more detailed analysis of citation stylistics between authors.

Results from an analysis of socio-cognitive proximity using citee and co-citation networks and their relationship to citing identities is presented. Findings from these analyses are discussed in relation to the use of the developed citing identity types, and the success of these methodologies discussed for the Consciousness authors.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

The findings of the research will be discussed in terms of the application and development of citing identities and theories, and recommendations for further work made to continue development in these areas.

2. CITATION THEORY AND THE SITUATED COMMUNICATING INDIVIDUAL IN KNOWLEDGE PRODUCTION

This chapter examines the potential afforded by the dominant theoretical positions on citation practice to contribute to furthering understanding of the representational outcomes individual scholars produce within the course of their scientific work practices. This examination will contribute to building on the foundations of citation identities in relation to citation theories and their future potential for studying the scientific work practices of individuals.

2.1. CITERS AS COMMUNICATING AUTHORS

Communication is central to the scientific work practices of an individual within knowledge production; individual scientists are integrated within a 'collectivity of scientists' from which they receive their 'subject matter', 'means of production' and 'target' through the collectivities common body of knowledge (Gläser, 2001, pp.193-194). One aspect of researchers' communicative and scientific work practices includes the production of publications and this involves writing. Through the writing process knowledge claims are developed or presented and 'linked' to the work of their colleagues (Amsterdamska & Leydesdorff, 1989; Borgman & Furner, 2002; Budd, 1999; Cozzens, 1989; Leydesdorff & Amsterdamska, 1990; Small, 1987; Small, 1998). Publications and their associated references are the outcomes of this communication process, becoming stabilised representations from the processes of iterative artefact and meaning construction (Lenoir, 1998; Small, 1978). Citers and therefore citation identities sit within the role of the individual knowledge producer as communicating author (White, 2001, 2004a).

In order to adequately accommodate for the role of the individual in knowledge production a citation theory needs to concurrently address both the scientific work practices of an individual as a producer of knowledge and allow for the placement of the individual within their communicative environment.

2.2. CITATION THEORIES

Citations in their simplest sense are the same as the references that are attached to scholarly works, providing points of referral to anyone reading that work (Langham, 1995). However in order to begin to utilise the abundance of data potentially available through citations, an understanding of what they could possibly mean or represent

needs to be obtained. Recognition of the possible significance of distinguishing between references (as allocated by the citing authors) and citations (as allocations to a source from another source) is established (Wouters, 1999b). However this is a distinction not often explicated in the field of citation studies allowing for ambiguity to arise (Egghe & Rousseau, 1990). Citations and references are again in the simplest sense the mirror image of one another; however Wouters (1999b, pp.10-12) argues, highlighting the many authors that go before him (starting with Price), that the difference is crucial to developing improved theoretical accounts of citation and reference behaviours. This argument is supported empirically from the range of heterogeneous findings that accompany a shift in perspective from one to the other (Gilbert & Woolgar, 1974), or when they are considered in relationship to one another (Amsterdamska & Leydesdorff, 1989; Baldi, 1998; Leydesdorff & Amsterdamska, 1990; Small, 2004b).

Citations have been characterised as representations of the choices scientists constantly make ‘regarding what contributions they should accept’, which in turn determines the ‘position of other scientists in the conversation of science’ (Baldi, 1998, p.830). When viewed in this pragmatic way citation analysis is a potentially rich source of empirical evidence for representational aspects of scholarly work practices and how these practice outcomes interrelate to other scholars’ communication outcomes. However no citation theory on its own supports this seemingly straight-forward pragmatic account, thereby providing the mechanisms that are needed to be understood to develop a theoretical basis for such an account in terms of a situated individual knowledge producer. The following review examines the theoretical literature on citations and references and asks the question of how the currently available theories contribute to understanding the role of the individual and their representations in knowledge production.

2.3. A TYPOLOGY OF CITATION THEORIES

Citations have been the focus of many and varied studies, ranging from the analysis of potential mechanisms involved in the practice itself (Case & Higgins, 2000; Chubin & Moitra, 1975; Shadish, Tolliver, Gray, & Gupta, 1995) to studies relying on citations as descriptive and evaluative measures of cognitive and social phenomena in knowledge production (Cole, 1975; Noyons, 2001; Noyons, Moed, & Luwel, 1999; Small, 1973, 1980; Small, 1999a, 1999b). Baldi (1998) notes that between 1965 and 1979 attempts to capture the reasons for citing work ‘became something of a cottage industry’ creating ‘no less than 10 different classification schemes’ (p.831). Referring to observations made by Kaplan (1965), Mulkay (1976a), Cozzens (1981) and Cronin (1981), Zuckerman

(1987) goes as far as saying 'by now, it may seem redundant to say that a theory of citation is badly needed' (p.336). The time following Zuckerman's assertion has not seen such a theory produced, which is still a prominent point of concern as evidenced by the findings and interpretations of recent accounts of citations and citing behaviour. The findings from these studies demonstrate that a clear empirical distinction between the two dominant positions within the theoretical literature is not easily obtainable (Baldi, 1998; Kurtz et al., 2005; Moed & Garfield, 2004; White, 2004b).

Due to the complexity and range of theoretical and behavioural positions available regarding citations and references, a typological approach is developed here that focuses on the prominent theoretical accounts that have implications for the use of citations in understanding individual researchers work practices. It is recognised that this typology is not comprehensive, particularly in regard to motivational and behavioural studies on citation, though it is considered indicative of the broader range of accounts.

Emphasis within this typology is placed on the analytical distinction between motivation and behaviour, as they appear to be manifested in citation theory literature. The distinction is noted as it has a defining role in interpretations and findings surrounding citation practices. The significance of the analytical disambiguation of motivation and behaviour is of course not restricted to citation theory and the complexity of their inter-relationship is noted. Ajzen (1991), for example, demonstrates the many factors involved with attempting to account for behavioural variance, such as intention, attitude to behaviour, subjective norms, past behaviour and perceptions of behavioural control. These factors are further complicated by variation in levels of aggregation. In the sociology of science Zuckerman (1988) notes the distinction between behavioural evidence of conformance to norms and actual conformity to those norms. Analytical differentiation between intentions, perceptions and attitudes, and behavioural outcomes has only been partially recognised within citation theory and has been argued to have contributed to limitations in developing citation theory at different levels of aggregation (van Raan, 1998).

The categories as developed here are not considered to be necessarily mutually exclusive, though where appropriate the significance of their placement is discussed.

The prominent citation theories have been categorised into the following:

- Attribution theories of citation
- Associative theories of citation

The attribution citation theories as presented in this typology include the tactical theories of citation and the normative theory of citation. Both of these theories attribute citation practice to an underlying form of authority, though the fundamental causes of the role of authority is conceptualised differently between them.

The tactical theory of citation is explicitly motivational; authors are motivated to cite certain works, as they convince the reader due to the perceived authority of the cited work. The success of the use of references to convince is determined by the use of the cited work in the argument as contributing to the support of the knowledge claims of the citing work.

The normative theory of citation is behavioural as it is embedded within Merton's institutional sociology of science. According to Merton's framework knowledge producers are socially regulated by institutional norms. When extended into citation theory, authors are expected by this theory to attribute credit, through acknowledging intellectual debt, influence, or use in developing the research work represented in the publication, and therefore recognise appropriate authority and priority according to the norms that guide the broad level values of science. In the normative theory of citation, references are attributed according to substantive use in the discovery process undertaken by the knowledge producer in establishing the research work as it is represented. The reward system is invoked when others use this work and therefore recognise its contribution to the new work.

A crucial defining feature between the two attribution theories is what constitutes the basis of recognition of authority in the cited source. The normative theory expects use of sources to be closely aligned with the original author's intention and content, whereas the tactical citation theories expect the use of sources to involve interpretation and deviation from consensus meanings.

The associative citation theories as presented in this typology include the symbolic theory of citation and the self-organising communication systems theory of citation. This category of citation theories include theories that take as their focus the communicative and epistemic functions of citation practice as they are embedded within broader knowledge production processes, networks or systems. Both theories are based in the association of citing and cited through meaning interpretations of the citer and cited work relationship.

2.4. THE NORMATIVE CITATION THEORY

Appearing in 1942 as a paper entitled “A note on science and democracy”, Robert Merton’s normative theory of science is the first attempt to provide a systematic conception of the social structure of science (Merton, 1973, p.xviii). This original paper (republished in (Merton, [1942] 1973)) provides Merton’s initial detailed discussion and definition of the four ‘norms’ of science: universalism, communism, organised skepticism, and disinterestedness.

Universalism (and counter-norm of particularism):

...truth-claims, whatever their source, are to be subjected to *pre-established impersonal criteria*: consonant with observation and with previously confirmed knowledge. The acceptance or rejection of claims entering the lists of science is not to depend on the personal or social attributes of their protagonist, his race, nationality, religion, class, and personal qualities are as such irrelevant ((Merton, [1942] 1973, p.270), Italics in original)

The norm of universalism is in direct opposition to the counter-norm of particularism. Particularism describes the imperative that non-objective criteria cannot be used to ascertain scientific validation (or allocation in the reward system); this also raises the dual issues of objectivity and relevance determination. The evidence of validation or invalidation must be pertinent to the scientific claim under examination, and the basis of this criterion must not take any extraneous non-scientific issues into account.

Zuckerman (1988) combines the original statement of the norm in Merton’s 1942 paper with a statement from the later 1957 paper (Merton, [1957] 1973, p.515), adding that ‘the norm of universalism also requires that scientists be rewarded in accord with the extent of their contributions to science’. Zuckerman also notes that in a variety of activities within scientific practice the norm of universalism is found to be applied in ‘neither wholly universalistic nor wholly particularistic’ ways and that particularism modifies decisions on contributions and receipt of rewards, and it is applied ‘along with role performance’ (Zuckerman, 1988, p. 518), following the more universalistic norm, neither norm is applied in an exclusive fashion. The analytical separation of the norm of universalism and its counter-norm is raised in the normative theory but in practice they are found to be integrated.

Communism

The norm of communism mainly concerns scientific ‘property rights’ and ‘common ownership’.

The substantive findings of science are a product of social collaboration and are assigned to the community. They constitute a common heritage in which the

equity of the individual producer is severely limited...The scientist's claim to "his" intellectual "property" is limited to that of recognition and esteem which, if the system functions with a modicum of efficiency, is roughly commensurate with the significance of the increments brought to the common fund of knowledge (Merton, [1942] 1973, p. 273)

The individual scientist only has rights to the esteem and recognition produced from the credit of the work given to them by the scientific community, but they do not own the right to exclusivity of use over the work, this second belongs to the scientific community. Zuckerman (1988) places another emphasis on this norm which is essentially translates into the imperative to publish; priority and recognition only follow from the public communication of work to scientific peers.

Disinterestedness

Disinterestedness is the 'curbing of personal bias' (Zuckerman, 1988, p. 515). However, Merton cautions against confusion of the institutional and motivational levels of analysis when discussing disinterestedness, stating that disinterestedness is 'is not to be equated with altruism nor interested action with egotism' rather it is 'a distinctive pattern of institutional control of a wide range of motives which characterises the behaviour of scientists' (Merton, [1942] 1973, p. 275-276).

Organised scepticism

The norm of organised scepticism calls for the critical and objective analysis of questions of fact. 'It is both a methodological and an institutional mandate' involving 'the temporary suspension of judgment and the detached scrutiny of beliefs in terms of empirical and logical criteria' (Merton, [1942] 1973, p. 277). Zuckerman highlights the institutional emphasis of this norm, 'for arrangements such as refereeing and other critical appraisals of work by competent peers: not necessarily for each scientist to feel uniformly skeptical' (Zuckerman, 1988, p. 515).²

Merton's norms (Merton, [1942] 1973) and his conceptualisation of the reward system in science (Merton, [1957] 1973, p.515) integrate to produce the foundations for Merton's institutional approach to understanding science as a social system. It is the interaction between the components of the postulated set of norms and the reward system that provides the explanatory mechanism driving behaviour in Merton's institutional framework.

² Further norms are developed after the original four for example: humility (begins p. 303) and originality discussed further in (Merton, [1957] 1973). See also (Barber, 1952) for a review of other values of science and a discussion of the later norm of rationality – "the critical approach to all the phenomena of human existence in the attempt to reduce them to ever more consistent, orderly, and generalized forms of understanding" (Merton, 1973, p. 225). However the CUDOs norms detailed here are the central ones for citation theory.

The concept of norm has a particular role within Merton's theory; this also has important implications for later developments and usages within citation theory development. Merton's theoretical complex moral and technical norms implement institutional imperatives (mores), that in turn support and derive from science's goal of 'the extension of certified knowledge' (Merton, [1942] 1973, p.270). 'Certified knowledge' is theoretically defined by science's technical methods, science is an 'empirically confirmed and logically consistent statements of regularities', and is supported (and defined) through the whole system of goal, imperatives and norms.

Merton's normative structure of science is explicitly developed as a 'value' system, and as an 'ethos' (Merton, [1942] 1973, p.268-269); it articulates the scientific motivational ideal and it describes scientists behaviour in an optimal or idealised form³. This is opposed to a specifically descriptive account of accumulated instances of strictly uniform behaviour or motivational imperatives. As such Merton does not claim that scientists adhere universally, comprehensively or exhaustively to the norms he puts forward; indeed he further develops his theoretical framework to include potential systematic deviance from these norms as well as the potential conflicting nature of the norms themselves. As one example, in an analysis of priority and recognition within the reward system of science, using the Darwin and Wallace priority dispute as his study, Merton highlights the potentially conflicting nature of institutional emphasis derived from norms, in this case the values of originality and of humility and modesty (Merton, [1957] 1973). Where Merton does explicitly discuss the evidentiary behaviour of scientists variation is a key theme; expected norms of the past are seen to be shifting with the advent of new social and cultural aspects of science. The particular examples Merton emphasises are the phenomena of multiple authorship, and that more generally 'scientists vary greatly in their attitudes to competition', including positing that Newton's reactions within his priority dispute over invention of calculus with Leibniz was due to the extreme emphasis that the 'newly institutionalized value set upon originality in science' had created (Merton, [1968] 1973).

Merton's normative definitions for the structure of scientific motivation provide a scaffold against which conditions leading to conformist and deviant behaviour can be examined. This scaffold crucially exists within a larger complex of sociological theory, where the social structure of science is driven by social institutions of science. The

³ Though Zuckerman in a response to Mulkey's (1976b) critique, that norms are essentially limited to ideological status, argues that norms are not 'just' ideological statements designed to defend the autonomy of science, they have far broader sociological implications despite the fact that empirical questions still remain about conformity to and deviance from them (Zuckerman, 1988, p.517)

institutional nature of science provides the legitimatisation of these norms and hence encourages their maintenance in the form of ‘prescriptions, preferences, and permissions’ (Merton, [1942] 1973, p.269), further emphasising the role of norms as prescriptive: what a scientist should do rather than necessarily a description of what they actually do. The reward system is the self-defining mechanism that imposes the constraints of conformance to institutional values, subscription to norms is driven by receipt of institutionally defined rewards; the alternative deviance from norms raises the possibility of institutionalised forms of ‘punishment’(Zuckerman, 1984).

Mitroff (1974) further develops the implications of the relationship between norms and counter-norms within science production, and maintains that the combination of both aspects contributes to science. Norms and counter-norms are features of the reward system of science, and contribute to the institutional framework as a whole; scientists are encouraged to comply with norms through this mechanism and its role in the centrally placed communication system of science (Storer, 1966).

Merton (1988) articulates a functional understanding of the role of referencing in the reward system of science,

As part of the intellectual property system of science and scholarship, references and citations serve two types of functions: instrumental cognitive functions and symbolic institutional functions. The first of these involves directing readers to the sources of knowledge that have been drawn upon in one’s work. This enables research-oriented readers, if they are so minded, to assess for themselves the knowledge claims (the ideas and findings) in the cited source...They also have not-so-latent symbolic functions. They maintain intellectual traditions and provide the peer recognition required for the effective working of science as a social activity. (p.621)

The normative theory of citation is generally acknowledged to first have been explicitly articulated by Kaplan (1965).⁴ Drawing on findings by Derek De Solla Price of ‘regularities’ and ‘discernable patterns’ evidenced in citation practice, Kaplan concludes that this suggests ‘certain norms’ driving citation behaviour. Through linking citation practice as the ‘coping’ mechanism of ‘property rights and priority claims’ via acknowledgement of intellectual indebtedness within Merton’s institutionalist framework, the connection between Merton’s normative sociology of science and citation practice is established.

Kaplan explicitly links Merton’s moral (or the later social) norms of Universalism, Disinterestedness and Communism, incorporating an understanding of the fundamental ‘value’ based nature of the Mertonian system and its articulation as an

⁴ Small (1998, p.143) dates it ‘at least’ this far back.

‘ethos’ which describes the scientific motivational ideal (Merton, [1942] 1973).

Recognising that little is known about the normative mechanism of citation practice and that there are many potential individual variations in the process, Kaplan then frames the remaining development of citation theory in terms of the discovery of the nature of the normative mechanisms that specifically drive citation practice. Kaplan’s establishment of the connection between Mertonian normative sociological theory and citation theory remains closely aligned with the sociological programme as presented by Merton ([1942] 1973).

In a review of Merton’s sociological theory of science Zuckerman (1988) draws together the literature that has followed its introduction, including the empirical studies that have been undertaken to examine scientists commitments to the norms. Zuckerman finds much of these analyses incomplete; those on the individual level are ‘sketchy’ (Zuckerman, 1988, p. 517) and in the aggregate variety is demonstrated but ‘firm conclusions would be premature’ (Zuckerman, 1988, p. 518).

Normative systems are never perfectly integrated and often call for conflicting behaviors in differing contexts. Individuals adapt to such conflicting normative expectations by oscillating between normatively legitimated but inconsistent behaviors, and in the process accommodate to the special demands of given situations (Zuckerman, 1988, p. 518).

These findings combined with Zuckerman’s statement about the normative system itself, raise the question as to whether it is possible to find evidence that disproves the theory. Conformity and deviance is accounted for only at the very broadest level, though detailed identification of the factors and conditions that surround compliance and deviance is not theoretically accounted for. The ‘normative’ nature, and the implied relationship between motivation and behaviour, of the theoretical framework itself hinders the development of these accounts. Merton himself highlights the problematic implications of this imprecise relationship between norms and behaviour,

how to account for regularities of social behavior that are not prescribed by cultural norms or that are even at odds with those norms. It casts doubts on the familiar assumption that uniformities of social behavior necessarily represent conformity to norms calling for that behavior (Merton, 1959, p. xxiii)(cited in (Merton, 1973)) .

This indicates that there not a clear relationship between norms and behaviours in Mertonian theory; therefore the sorts of scrutiny and expectations placed on the extrapolation of normative accounts into citation theory is similarly subject to these constraints. Studies that attempt to disambiguate normative and tactical citation theories often ascribe positive behavioural findings to conformance to norms (for

example (Baldi, 1998; Kurtz et al., 2005)). Concurrently what may be cases of deviance are analytically distinct yet empirically difficult to discern from tactical theories of citation.

The structural and institutional basis of Mertonian theory is continually emphasised within the framework and has implications for the placement of individual scientists. The original normative theory ‘asserts the ethos is “procedurally efficient” at the institutional level and makes no claim that this should hold for individuals’ (Zuckerman, 1988, p.519). Small (2004a) relates this perspective to citation theory reminding us that the Mertonian system is a value system and therefore a citation norm, like other norms, is not a rigid constraint on behaviour but is a ‘set of expectations for conventional scientific behaviour against which we can measure deviations’ (p.72). From these accounts it can be seen that agency of individuals is not addressed within normative theory, and this follows into its citation counterpart.

Significant development of the normative citation theory subsequent to its establishment by Kaplan and Merton has been based in its juxtaposition with the tactical theories of citation and methodological critiques of citation analysis. These aspects of theory development are discussed in Section 2.6.

2.5. THE TACTICAL CITATION THEORY

Gilbert’s (1977) articulation of a tactical theory of citation is in direct and specific response to Kaplan’s citations as intellectual property rights argument. Gilbert notes that Ravetz has two arguments; the first is the same as Kaplan’s, while the second is identified as citations as they are used to construct an argument (pp.114-115). Gilbert then proceeds to argue against the ‘norm’ of intellectual property, using empirical evidence provided by perfunctory and negational citations, and argues for its substitution with a more refined version of Ravetz’s second function of citation (p.116). Gilbert develops Ravetz’s brief outline into his persuasion hypothesis. The central tenet of the hypothesis is that citations are one of a number of rhetorical devices used by authors to enhance the persuasiveness in their arguments for knowledge claims. This persuasion is entrenched within the previous work of the scientific community that has ‘already been accepted as “valid science”’, and therefore gains its ability to persuasively support the new claims (p.116).

In order to justify an argument to an audience of potentially interested readers, it is most effective to cite a selection of those papers which the intended audience believe present well founded, valid results (p.116)

and again including the concept of audience dependent importance:

In order to support their research findings, authors will tend whenever possible to cite papers which they consider their audience will regard as presenting valid important arguments and results (p.118)

This involves a postulated 'shared belief' or some degree of consensus on what will constitute 'important and correct' as well as 'erroneous', 'trivial' and 'irrelevant'; all of these evaluative positions are determined by the scientific community/ies that comprise the 'audience'. Therefore it is expected within Gilbert's account that 'authors preparing papers will tend to cite the "important and correct" papers' but may also legitimately 'cite "erroneous" papers in order to challenge them' and will avoid the 'trivial' and 'irrelevant' ones (p.116).⁵ Through this selection an author provides support for their work and simultaneously is displaying 'allegiance to a particular section of the scientific community' (p.117). Exemplars arise from the works' repeated use as authoritative grounds for further work and so can be used to indicate the consensus within scientific communities. Gilbert supports his hypothesis with how it can be used to explain empirical evidence by including the works of 'the Coles' and Small and Griffith (1974)⁶:

Their technique is successful because authors, in choosing references (and thus co-citation pairs) orient to their own perceptions of how the scientific community and its knowledge is structured. They place their work within a field by citing research which their intended audience values. Thus the co-citation analysis reveals the specialty structure by jointly tapping the individual perceptions of all the authors whose work has been examined. (pp. 118-119)

Gilbert's hypothesis accounts for perfunctory and negational references and their function in supporting knowledge claims. This support is crucially contingent upon and directed toward 'the interests and knowledge of a particular audience as they exist at a particular point in time' (p.119). This position argues against the ability of non-specialist readers to determine the precise functional nature of the reference according to criteria of 'perfunctory references' as put forward by citation context analysis by Chubin and Moitra (1975) and Moravcsik and Murugesan (1975). This form of analysis requires that the reader ascribe 'intention' on the citing authors usage of a reference; however this is problematic from any external reading of intention, more particularly

⁵ This is hardly a haphazard approach to citing; see (Peters, Braam, & Van Raan, 1995, p.21) 'arbitrariness mainly for adornment of claims' (though Peters only refers to interpretation of Cozzens in the text – but attributes this to 'some circles of sociologists of science' more generally

⁶ It is noted however that Small maintains the applicability of Kuhn's paradigms as foundational within his work, though modifies his interpretation over time (see (Small, 2003)), and Gilbert specifically argues against the 'incommensurable' shifts required by Kuhn (see (Gilbert, 1976)). This is important to note as it contributes to the two researchers different contingent interpretations of the fundamental processes of scientific change.

readings as they are interpreted by a non-member of the appropriate scientific community (pp.119-120).

References and citations are central to the processes of generation and evaluation of knowledge claims within Gilbert's hypothesis. Though Gilbert places his focus in 'the context of justification', in order to distinguish it from other "'micro"-scale studies' that are concerned with the 'context of discovery' (Gilbert, 1976, p.281) he states that the processes are empirically indistinguishable (p.288). In this process the 'only public evaluation' of knowledge claims 'occurs when it is cited as supporting evidence for further knowledge claims' (Gilbert, 1976, p.296). Myers' (1993) study on the shaping of two biology articles, presents evidence that supports and extends aspects of Gilbert's persuasion hypothesis, throughout the process of writing, formal and informal reviewing and publication. Myers' focus is on the 'status' of the knowledge claim and how this is negotiated through the writing and publication process. References play a crucial role in establishing knowledge claims, and comments of formal reviewers reflect acknowledgement of this role in their comments specifically related to inclusion and exclusion of appropriate literature. Myers defines the hierarchy of knowledge claims according to the 'distance between the authors' claims and the claims of the particular part of the scientific literature in which it is placed' (p.335). One of the biology articles is determined to be 'speculative' not 'because it runs ahead of the data, but because it runs ahead of the literature' (p.340). Myers extends conceptualisation of the persuasive and rhetorical role of references to communicate knowledge claims, providing evidence that they are integrated into the acceptance of the knowledge claims represented in the publications as well as their status. Though specially excluded by Gilbert (1976; 1977), Myers introduces the review process, journal selection, review and initial publication acceptance as playing a highly significant role in determining how claims outcomes are presented to scientific audiences.

Science for Latour (1987) is a fact construction process. This process involves two sorts of facts: those 'in the making' and those 'ready made' (p.4). However the differentiation between these two kinds of facts is only time and situation dependant; the same fact is one or the other as placed in time and context. Facts exist as two sides of the same coin, or to use Latour's metaphor 'the two heads of Janus'. The status of facts is continually being reassessed through the course of their use in scientific work. This work either redefines the facts' status as subject to question and therefore reverting it to 'science in the making', or the fact maintains its status as part of 'ready made science'. The transition from the first to the second creates a fact 'blackbox', and is used as a basis to

establish further facts. The role of fact construction and its Janus-like qualities is central to Latour's theory of citation. Citations have the role of contributing to the establishment of scientific facts, both through the justification and communication process and also through the decisions upon which their use is based.

For Latour, previous literature is one of the resources scientists use to construct, place and justify their claims. Decisions on the use of that literature involve assessment of their status as facts, and their specific contribution to the fact making of the claim. Use and placement of that previous literature happens both during the research process and within the justification process; these processes are indistinguishable in Latour's theory. Decisions of the use of previous literature happen continually during fact making, but also crucially for his citation theory, within the writing of publications. Citations are part of the enrolment and rhetoric process and the rallying of convinced 'allies' upon which further use by scientists is dependant. This further use is the crucial mechanism for the establishment of the transition of the "fact in the making" to the "ready made fact"; publication is only one of the earlier hurdles for the acceptance and establishment of a ready made fact.

Reliance on previous literature as a convincing resource means that the resource must be recognisable as convincing within the construction of the knowledge claim, and therefore its role in the knowledge claim and its associated status is communicated in the course of the argument. Further integration and use is vital for the fact to become and remain ready made; this also involves the fact having to be used positively as a contribution. Continued negative or disconfirming citations will lead to the fact's eventual removal from both its status as fact but also its role in science (p.40-41). Positive modalities (positive use of work in further work) moves the fact away from conditions of production (toward ready made science/established fact) whereas negative modalities (countering use) move the fact toward conditions of production (toward science in the making) (p.23).

Latour accounts for the sociocognitive aspects of references-citations in terms of fact use, fortification, persuasion and acceptance by other scientists. Latour's theory accounts for the diversity found in empirical studies of citation function such as (Amsterdamska & Leydesdorff, 1989; Case & Higgins, 2000; Chubin & Moitra, 1975; Leydesdorff & Amsterdamska, 1990; Moravcsik & Murugesan, 1975). The ability of Latour's theory to account for the empirical heterogeneity found in citation function is also noted by Luukkonen (1997); however to some extent this alignment of findings is not surprising given Latour's theory of citation is in part also developed from the

context of citation methodology. Latour's theory of citation also accounts for the diversity found in motivational studies (Brooks, 1985, 1986; Wang & Soergel, 1998; Wang & White, 1999; White & Wang, 1997), though it should be recognised that in this analysis Latour is ascribing motivation to behavioural outcomes as assessed through his citation context study of the GHRH/GRF controversy.

Latour's citation theory is centrally involved with the enrolment of 'allies', those that are convinced to accept and use the authors work; however Latour only provides an account of 'allies' in terms of them as other individual knowledge claim makers. Positioning and success of the claim is determined by the shifting balance between 'isolated' communicating actors. An author or a reader is 'isolated' when they are not supported by sociocognitive connections; the actor least supported by connections is likely to be either convinced (reader) or unconvincing (author) (p.50). In the course of his analysis Latour for the most part compares the authoring scientist with the 'average' reader; the average reader is convinced through support garnered in arguments from authoritative and placed previous literature, and the work represented in the publication via argument and inscription devices. However it is other scientists that have the ability to collectively translate the fact in the making to the fact ready made. As for the average reader, the same convincing is, to some extent, also needed for another scientist, but unlike the 'average' reader, another scientist is not as likely to be as easily 'isolated'; they themselves have 'fortifications' from their own work and the networks of scientists in which they are members. It is the coordinating function of 'groups' and institutional actors that is not addressed in Latour's theory of citation. These 'groups' and their participation in constituting the institutional actors within knowledge production, function at various levels throughout knowledge production processes.

Whitley (1980) discusses the significance of institutional mechanisms, using among other examples, the role of journals in organising science and therefore the significance of publication acceptance (pp.315-316). For a fact in the making to even be considered within the public conversation of science, and therefore in Latour's terms progress toward becoming a ready made fact, it needs to first be published (Zuckerman & Merton, 1971). Latour (1987, p.45) recognises the significance of the role of the journal in the 'hierarchy' of science, when he states that the status of the journal contributes to the reception of the fact in the mind of the reader (p.45); however journal audiences are not only constituted by the 'average' reader, they are primarily constituted by other scientists and include scientists that work in/belong to membership groups which entail social control mechanisms (Whitley, 1984, p.28) and problem areas and

networks which contribute to defining span of control for influential groups such as the scientific 'elite' (Mulkey, Gilbert, & Woolgar, 1975). Scientists will bring their own meanings and interpretations based on their own work, and the meanings they have generated through the course of producing that work, in relation to their research networks (Gilbert, 1976). Small (2004a) differentiates 'constructive' with 'normative' accounts of citation practice according to the development of meaning consensus (p.76-77). Latour does account for the reader that will bring their own meanings and interpretations, in his theory of fact making and citation, but what are not accounted for are the levels of consensus developed through the communication mechanisms and appropriate scientific work practices as primarily established at institutional levels.

Latour (1987) says interpretations will be determined by readers and that those readers have a context, but he does not account for the context derived from intermediating institutional groups, such as research traditions (Nicolaisen, 2003, p.18) and specialties (Gläser, 2001), which is important for understanding citation practice and individual scientists work practices. If nothing else, initial publication acceptance is dependant on these institutional levels, and as Myers (1993) empirically demonstrates, reviewers have a role in accepting the status of a claim in relation to broader networks of scientists, the claim's presentation and placement in a hierarchy of journals and the references, and arguments that are used to support the claim. Gläser (2001) highlights the significance of institutional mechanisms in the knowledge production process; they create the social order in which knowledge is produced, and the possibility of studying specialties provides an important link between sociological and citation studies of science. Latour follows the collective establishment of facts from fact in the making to fact made, but the acceptance of the fact is only accounted for as generally scientific, rather than its placement within groups determined at much more refined levels of institutional hierarchies of acceptance and fact establishment. In this regard Latour's theory of citation does not adequately account for the situated role of the individual in knowledge production. In terms of an individual's citation practice his focus is on the individual as situated, but only in regard to the very broadest levels of science, not in their positioning in relation to their institutional affiliations.⁷

⁷ Luukkonen (1997) in her discussion of the potential relationship between Small's concept symbols and actor-network theory, notes the parallels in Callon, Law and Rip's use of actor network theory in their application of co-word analysis, and believes it may imply that co-citations could usefully be considered as associations within 'networks of problematisations' (p.33). Introduction of Callon's (1980) identification of 'problematic situations' may indeed contribute to further development of Latour's theory and its relationship to scientific groups or institutions; however Luukkonen herself only introduces the possibility of this convergence, and

2.5.1.1. AUTHORITY AND USE IN ATTRIBUTION THEORIES OF CITATION

Regardless of the possibility of heterogeneous modalities within which a reference is allocated, Latour's citation theory does involve critical use. In both tactical theories of citation, though their use is heterogeneous, citations have a unifying function in regard to presenting and developing knowledge claims; that is, to support knowledge claims (Latour, 1987, p.38) and to justify argument through a selection of 'papers which the intended audience believe present well founded, valid results' (Gilbert, 1977, p.116). The rhetorical functions of citation involve more than simple appeals to authority; these appeals are embedded in the collective and accumulative authority attributed to the work the citation represents, and the further assessment by readers of the degree of attachment of the reference to the claim is it used to support (Latour, 1987, p.33). Authority is an important aspect of the two tactical theories, as it determines the success of the main function ascribed to them.

Authority and its definition is also a key issue in normative citation theory, through the question of what makes a source influential. Zuckerman (1987) specifically refers to the difficulties inherent in defining normative citation theory as opposed to tactical theories of citation with regard to 'authoritative' use of references. She expects there to be 'overlaps' in the behavioural outcomes stemming from attribution with the intent to persuade and 'legitimate' attribution according to intellectual influence (p.334). Small (1998) also indicates the difficulties associated with some forms of behavioural empirical testing of the attribution theories: 'both theories seem to predict that scientists are more likely to cite highly cited authors than low cited authors' (p.143). These definitional problems transfer into their comparative operationalisation, which is discussed in the following section.

2.6. EMPIRICAL TESTS OF THE ATTRIBUTION CITATION THEORIES

The attribution theories of citation, the normative theory and the tactical (social constructivist or micro-constructivist) theory, form the focal point and dominant positions for much of the development of citation theory. Historically they are the foundational approaches and still form the basis of the majority of current theoretical discussion within the field, either for the purposes of extension (eg (Baldi, 1998; Cronin, 1984, 2004; Small, 2004a)), support (eg (Kurtz et al., 2005; Moed & Garfield,

this has not been further taken up and developed into empirical and theoretical accounts of citation.

2004; White, 2004b)) or points of departure (eg (Cozzens, 1989; Nicolaisen, 2003; Wouters, 1999a)).

The majority of contemporary citation theory maintains the long-standing centrality of the attribution theories, through the argument that understanding the nature of why citations are assigned (or more precisely referenced) will contribute to the application of citation analysis and will provide the logical limits of the role of citation in understanding scientific practices (Bavelas, 1978; Cronin, 1981).

Significant emphasis in the further interpretation and development of normative citation theory, beyond the earlier formulations of Merton and Kaplan, has been placed in addressing critiques raised by the tactical theories of citation; as the tactical theory invokes intentional assumptions, many attempts have been made to address the comparative testing of the theories in this regard. However it becomes clear when referring back to Merton's work (see section 2.4) and to Small's (2004a) use of the concept of 'symbolic payment of intellectual debt' (p.76), that normative citation theory is a structural-functional theory of citation, and that reconciling the structural-functional nature of the theory with motivational interpretations is difficult. Borgman (2002) describes the 'interpretativist and structural trends' within citation studies as oppositional. Interpretativist theory (tactical citation theory) emphasises the citer's personal actions, 'influenced but not determined by context', whereas structuralist-functional (normative citation theory) approaches emphasise concern with 'identifying probabilistic regularities and patterns' in behaviour (p.46). Therefore the motivational implications of the normative citation theory have necessitated re-interpretation from the Merton's sociology of science, and many attempts have been made to produce satisfactory accounts of these mechanisms, and to operationalise them successfully.⁸

2.6.1. OPERATIONALISATIONS OF THE ATTRIBUTION THEORIES OF CITATION AND CONVERGENT FINDINGS

An example demonstrating the difficulties associated with operationalising the theories comparatively, even when specifically attempting to examine intentional aspects, can be seen in Cronin's (1982) survey questions and definitions. Cronin defines normative citation theory in relation to its functionalist sociological foundations and offsets it with 'microsociological' accounts and tactical citation theory (p.52-53). However when constructing a survey to determine the significance of functions and norms in citation

⁸ Which was in fact what Kaplan (1965) recommended but did not himself pursue.

practice, these are separated according to lines not easily recognisable from these definitions (Cronin, 1982). Questions relating to norms are based in terms of conventions, 'established practice' and standardisation while questions relating to function concern tactical aspects, 'intellectual indebtedness', 'recognition' as well as 'symbols' (p.71).

The fundamental difficulty of operationalising comparisons of the attribution theories can be exemplified by Cronin's (1982, pp.70-71) second survey question 'authors tend to cite those whose views tend to support their own'. This question is intended to differentiate between subjective and objective motivations of citation along the definitions of the attribution theories given by Cronin. However the survey is directed at psychology journal editors and so they are making an assessment of motivation from their readings of the papers they have reviewed generally. From this perspective it would be difficult to disambiguate the use of citations according to what that support constitutes. According to the tactical theories of citation, authors are expected to support their work with citations and according to normative theoretical interpretations such as Zuckerman's (1987) this support will also be evident through ascribing the same outcomes to the process of attributing intellectual influence and priority. Subjective and objective motivations that determine either theory cannot be judged from this perspective. Support for this as a problem of operationalisation is suggested in Cronin's findings that the interpretative split is not evident in his respondents' replies (p.75).

Difficulties in developing operationalisations of interpretative splits between the attribution theories are also presented in studies using the concept of authority as a base. Moed and Garfield (2004) attempt such an analysis. They define the normative theory of citation as stating that 'scientists give credit where credit is due' and this definition provides justification for citations to be used as approximate indicators 'to trace intellectual or cognitive influence' (p.295-296). In opposition the constructive view 'takes the position that scientists cite to advance their interests, defend their claims against attack, convince others, and gain a dominant position in their scientific community' (p.296) This establishment of definitions firstly requires imposition of simplified distinctions between the two attribution theories, but also requires that these distinctions are able to be operationalised effectively.

Moed and Garfield (2004) utilise Zuckerman's (1987) assertion that if authors used citations according to the persuasion argument, citing distributions would contain more than 6% of 'authoritative' papers over the cumulated Science Citation Index

(SCI). From this assertion they operationalise 'authoritative' papers as those cited within the 10% most frequently cited within their research field. Research field is defined very broadly in this study being 'aggregates of journal categories' including 'Molecular Biology and Biochemistry', 'Physics and Astronomy', 'Applied Physics and Chemistry' and 'Engineering' (p.298). They find variation in use of 'authoritative' papers across these research fields, with 'basic' fields using more authoritative papers than 'applied' fields. However overall the use of 'authoritative' papers is found to range from 26% to 39% in the fields. More interestingly they find evidence that for the field of 'Molecular Biology and Biochemistry', 'authoritative' papers are excluded as reference lists get shorter. This is interpreted as evidence for selection decisions based on cognitive relevance, as authors drop 'authoritative' papers more readily than other types. Moed and Garfield recognise that they are using a somewhat arbitrary threshold for identification of 'authoritative' papers in this study, but beyond that they are also using extremely broad research field categories to be attempting to determine relevance.⁹ They are unable to provide support for one attribution theory as opposed to the other, and conclude that 'bibliographies do at least partly reflect authoritativeness as suggested by Gilbert' (p.303).

The question remains whether it is possible for 'authoritative' papers to be defined in this way, Gilbert's persuasion hypothesis does include that to be successfully persuasive the reference would have to be recognised by the papers 'audience' as important and this is contextually defined. Therefore it is unlikely that an analysis at such broad levels of aggregation would capture this. The analysis does leave open the alternative interpretation that papers with longer reference lists are more likely to be 'review' papers where it would be expected that 'foundational' (of broader significance) references would be more likely to be included. Myer's (1993) work suggests that status negotiation in the review and acceptance of publications includes the length of publication being negotiated in relation to the perceived significance of the knowledge claims presented. This could mean that longer papers are more likely to have a broader scope, and therefore would include more papers of significance at the broad levels Moed and Garfield are identifying.

Use is also a problematic basis of discrimination for attribution theories of citation when operationalised using interpretations that remove their discriminatory power.

⁹ There is a substantial body of literature from information science that requires consideration of significantly more specificity in accounts of relevance (see for example, for general contextual accounts (Harter, 1992; Howard, 1994; Schamber, Eisenberg, & Nilan, 1990) and (Anderson, 2005) for relevance as established in the research process).

Kurtz *et al* (2005) define the normative assumption as “the number of times a document is cited...reflects how much it has been used” (citing (Liu, 1993)). This interpretation clearly obscures much of the complexities involved in the possibility of juxtaposing the normative citation theory in terms of any other theoretical position available. Kurtz *et al* do not define their interpretation of the tactical theories but cite the MacRoberts’ (MacRoberts & MacRoberts, 1989) critique of citation analysis as the oppositional argument to the normative assumption. In effect they interpret their otherwise impressive results of close relationships between reading of articles and citing of articles by individuals as ‘prov[ing] that the normative theory of citation is true in the mean’ (Kurtz *et al.*, 2005, p.111). ‘Reads’ are operationalised through individuals accessing of either abstract, references only or the full-text in the NASA Astrophysics Data System through log files, and this is taken to represent use. Kurtz *et al*’s findings do indicate that authors ‘read’ the articles they cite, and do so before they cite; however it is difficult to see how this supports a ‘proof’ of the normative theory of citation in any more than the sense of empirically opposing one aspect of the MacRoberts’ primarily methodological critique. Kurtz *et al*’s interpretation demonstrates the problems encountered when normative citation theory is extrapolated into simplified interpretations and then further extrapolated into operationalisations according to those interpretations.

Hargens (2000) is a relatively recent use of a normatively derived structural-functional framework to examine citation patterns, and empirically determines that the implications of the attribution theories of citation represent the necessity for some aspects of convergence to be recognised. In this work Hargens returns to the study of reference networks in an examination of how authors use each others work. His study is at the level of the research area (seven research areas from a variety of disciplines spanning ‘hard’ to ‘soft’) and combines structural aspects of the research areas using reference network analysis with citation-context analysis. Hargens’ findings demonstrate marked differences in use of literature among the research areas: these results are not due to network density, network size and literature growth rates.

Hargens then examines the possibility that these different structures may in fact be due to variation across research areas in how researchers use each others work, particularly according to the differences in consensus of the relative importance of topics in the individual areas. The citation-context analysis done by Hargens relates the use of the literature to the functional needs of communication within the fields brought about by their structures. For instance, fields that have less consensus on the basis of importance

of research topic are found to draw more regularly on 'foundational' literature and use this literature more conceptually ('cite it for its thematic content' (Hargens, 2000, p.859) to establish the significance of their topic. Fields that have more consensus rather focus on the findings of others, work, assuming the importance of the topic is established and so rely more heavily on 'research front' literature. This finding also suggests to Hargens that use of literature also has a legitimate rhetorical and communicative function that goes beyond simplistic 'persuasion', that can be related to the structure and features of the research area to which researchers belong. 'I suggest that authors' use of the literature varies depending on the tasks they must carry out to write papers that colleagues will see as significant contributions' (Hargens, 2000).

2.6.1.1. "WHO YOU ARE" VERSUS "WHAT YOU DO"?

Attempts have been made to determine the primacy of ascriptive process versus achievement process in the allocation of citations. This involves delineating the operationalised variables according to one or other process.

Stewart (1983) examines influence and recognition (which he uses interchangeably) through delineating what he considers their constitutive components, achievement versus ascriptive processes. He traces these components back to Merton via Cole and the norm of universalism, and firmly places his work in the 'functionalist perspective'.

To the extent that influence is distributed on the basis of *what one says*, then we have universalism or an achievement process, but if influence is determined by *who one is*, then we have a Mathew Effect, or an ascription process' (Stewart, 1983, p.168).

Stewart's study explicitly assumes that both elements will contribute to scientific process however believes their relative importance is the key determining question. Focus is placed on intellectual factors, to account for the limitations put forward by constructivists, and is examined through variables that are considered potential determinants of citations to articles.

Stewart holds that there is a fundamental limitation in examining intellectual influence on the level of the individual, as it does not allow for adequate assessment of the 'effect of the specific characteristics of the individual articles' (Stewart, 1983, p.169). These influence variables that relate the individual to their articles are used to operationalise achievement (article characteristics) versus ascriptive (author characteristics) processes. Author characteristics, which reflect 'scientific accomplishment or experience' (Stewart, 1983, p.176) include previous productivity, professional age and

average quality of publications as measured by average citations to article. Article characteristics include article length, publishing journal and articles relevance. As relevance becomes important in the finding this variable is further described; determination of relevance includes a 'subjective assessment of the articles relevance based on Stewart's understanding of the knowledge area, and three measures based on relationships between the articles' references and the references in major plate tectonics articles (Stewart, 1983, p.175). Stewart's definitions of variables lead to the question of whether an author, their reputation and their work can be so neatly separated from one another.

Stewart finds that 'achievement processes are more important than ascriptive processes' (Stewart, 1983, p.175) in the allocation of citations for both plate tectonic and geology cases. This finding leads Stewart to the conclusion that 'universalism was most important in the allocation of recognition for scientific contributions' although 'some evidence for the Matthew Effect was shown by the significant effects of some author characteristics' (Stewart, 1983, p.185). However it can also be seen that Stewart's definition of 'relevance' plays a significant role in this finding, therefore 'relevance' or some unarticulated definition of content relatedness is implied by (necessary within) the normative theory of citation.¹⁰

Following Stewart, Baldi (1998) also examines processes of ascription and achievement in the allocation of citations. He extends work done by Stewart by including the 'dyadic' relationship between citing and cited articles. Interpretation of ascriptive and achievement processes are also modified. In Baldi's analysis ascription, or '*who one is*', is 'indicated by functionally irrelevant author characteristics, such as one's eminence...' (p.833). Conversely achievement processes, or '*what one says*', is indicated by the 'cognitive, perceived quality, methodological, or topical content of one's article' (p.833). While Stewart contains his conceptualisation of these processes within the Mertonian functionalist framework, Baldi considers ascription processes to be representative of social constructivist theory of citation and achievement process to be aligned with normative theory of citation. Concurrently Baldi acknowledges that the two processes 'are not necessarily mutually exclusive' (p.833). Findings from this study are interpreted as supporting 'a normative interpretation of citation use' (p.834); however Baldi further tempers this with an overall interpretation that supports 'a model of citation behaviour in which citations are simultaneously determined by the characteristics of citing and cited articles and their authors' (p.843). This final

¹⁰ It should be noted here that within information science relevance is itself a very complex notion (see (Anderson, 2004; Harter, 1992; Schamber et al., 1990)).

interpretation seems to indicate that author characteristics may not in fact be decidedly 'functionally irrelevant', and may not solely represent constructivist citation theory.

When the Mertonian sociology of science is translated into the normative citation theory in this operationalisation process, the problems associated with divisions between motivational and functional/behavioural accounts become more evident. An example can be seen in Baldi's (1998) essentially behavioural variables which are ascribed motivational interpretations in his conclusion. His results accordingly do not demonstrate the expected clear distinction between the tactical and normative theories of citation. This underlying difficulty in Baldi's analysis is recognised by Small (1998, p.143).

Both of these studies (Baldi, 1998; Stewart, 1983) utilise conceptualisations of an empirical distinction between ascriptive and achievement processes. In order to assess the primacy of one or other of these processes the analytical distinction must be operationalised; even so, both articles conclude that findings indicate that both processes are at play. This in turn opens the question as to whether the distinction is analytically useful and empirically examinable. However, by both accounts content seems to be an important factor in allocation of citations.

2.6.1.2. OPERATIONALISATIONS OF THE ATTRIBUTION THEORIES OF CITATION AND THE VALIDATION OF CITATION ANALYSIS

One of the early calls for a theory of citing was put forward by Mulkay (1976a), who recognised that his own study used citation patterns 'as an index of lines of intellectual influence' and that this use assumed an 'implicit theory of citing'¹¹. Mulkay concurrently recognises the methodological utility of using citations in this way as 'they cannot be distorted by the selective perceptions of participants' and so are 'relatively objective data' (p.111). Even during this early stage of the discussion of citing theory it can be seen that a number of criteria for appropriate methodological considerations for the application of citation analysis are established. Citations are required to be 'objective' and representative of 'intellectual influence'; 'quality', in most cases should be 'positive', be content driven and have expected integral functions within reward

¹¹ This itself was not new as Cole (Cole, 1970, p.281) also noted this assumption; however it was stated rather than critically examined in his work (Small, 1998, pp.143-144). Note that Mulkay is actually talking about 'citing' here rather than citations, and is directly opposed to the argument put forward by van Raan (1998), and highlights the necessity of distinguishing between references and citations first articulated by Price. Failure to do so adequately contributes to the difficulties in developing citation / citing theories (Wouters, 1997).

systems of science. Thus in these forms the normative citing theory is interpreted to be the foundational position in establishing citation analysis as a valid methodology.

A dominant feature of the evolution of the attribution citation theories is that much of the discussion and empirical analysis that surrounds them is via their direct juxtaposition. Small (1987, p.339) refers to this phenomenon by highlighting the ‘caricature’ of “references as influence” introduced by David Edge and continued in the MacRoberts’ work. Through this process associative allegiances have been formed which produce certain dichotomous positions, which are influencing the development of citation theories and their ability to address understanding of scientific work practices.

The two dominant attribution citation theories are based in two fundamentally different viewpoints of citation practice. Citations for Zuckerman (1988) become valid measures because they are an aggregated measure, and represent the accumulated collective decisions of many scientists; it when they are examined in aggregate that citations are not subject to particularistic influences. This attribute of aggregation carries with it a certain internally logical consistent validity for the purposes to which it is put by Zuckerman and colleagues, but this is an argument that is not transferable into the individualistic micro-process of citation. Reliance on aggregation as being crucial to the validity of citation analysis and theory is also argued by Small (1987, p.339) and van Raan (1998).

Studies and theoretical positions that support a tactical theory of citation do not address aggregation; rather they develop a theory of citation through micro-analysis of citation practice as demonstrated in laboratory studies (Latour & Woolgar, 1979), discourse analysis (Myers, 1993), and citation context studies (Gilbert, 1977). In response to the issues raised in tactical accounts, arguments have been put forward that theories of citation must be tested according to motivations, as particularistic motivations of the citing authors invalidate the objectivity of citation measures (Bavelas, 1978). Furthermore, adequate theories of citation cannot be further developed without understanding citer motivations (Brooks, 1985, 1986). The prevalent assumption that understanding of motivation will clarify the use of citations as an analytical tool has been questioned; the basis of their refutation hinges on the differentiation of citing versus citation practice (van Raan, 1998; Wouters, 1999b) and the problematic relationship between motivation and behaviour (Langham, 1995; Zuckerman, 1988).

The normative citation theory is focused on and gains validation from the use of aggregated measures, whereas the tactical citation theories are concerned with individualistic and localised practices in science. This results in problems in relation to testing between the two theories. Neither attribution citation theory successfully accounts for findings of micro-studies simultaneously with those produced in functional-structuralist accounts, and so do not allow for complete examination of the situated role of the individual and their scientific work practices that involve simultaneously content and rhetorically significant communication.

2.6.1.3. THE QUESTION OF EMPIRICAL DIFFERENTIATION BETWEEN INTERPRETATIONS OF THE ATTRIBUTION THEORIES

This section [2.6] has highlighted a number of empirical studies that have found it difficult to successfully operationalise the divisions implied by interpretations of normative and tactical theories of citation, and produce conclusive findings in support of one to the exclusion of the other. This problem rests on two fundamental concerns. Firstly there is the problematic relationship between motivation and behaviour and therefore what is the appropriate way to support understanding of the theoretical position of citation analysis. Secondly there is the problem of empirically and satisfactorily discerning one attribution position from the other.

The differentiation between the two theories remains a difficult task, as White (2004b) explains,

one must show that, in order to sway opinion, an author has knowingly cited a work by a big name (or a person in power, or a colleague, or a favorite) that is less relevant to the text than a work by someone not in these categories (p.116)

The difficulty of this task demonstrates the question of the empirical separation of the motivational interpretations of these theories as being methodologically problematic. This is further highlighted by numerous citation motivational studies that suggest that citing involves simultaneous normative and egotistical values (Borgman & Furner, 2002, pp.48-49).

Small (1998) highlights the problem in reference to the differentiation between 'intellectual influence' and 'persuasion',

direct empirical test of the two theories seems difficult, and we need to take a step back and view these two theories in a broader context...on the normative theory scientists cite highly cited authors because they provide the best sources, and on the constructivist theory because citing highly cited authors is more persuasive than citing low cited authors (p.143)

Zuckerman (1987) further argues that citation practice studies need to be concerned not only with authors 'intentions', but it is also necessary to address 'audience response' (pp.334-335). Isolated attribution citation theories are not able to gain an appropriately comprehensive understanding of the relationships of these factors within scientific work practices.

The difficult empirical separation between the motivational aspects of citing practice is amplified with the problematic relationship between motivation and behaviour, this relationship ultimately asks the question: in what ways will motivational approaches contribute to citation theory explication, and indicate that both need to be considered? What is made evident in the foregoing account of the attribution theories of citation is that neither position adequately accounts concurrently for the distributional regularities found in citation patterns and the heterogeneity found in citing practices of authors.

2.7. THEORETICAL CONVERGENCE BETWEEN THE ATTRIBUTION THEORIES OF CITATION

Following an examination of the foundational sociological theories and supporting empirical evidence, Cozzens (1989) concludes that 'citations stand at the intersection of two systems, a rhetorical system and a reward system; they reflect both at once' (p.438). Echoing Zuckerman's (1987) identification of the problem that these two systems are found to be 'analytically distinct' yet 'concretely indistinguishable: they both present as impetus and constraint in any given act of citation' (p.440). A synthesis of the two systems (normative and tactical citation theories) is then presented in the form of a rhetoric-first model of citation practice. The rhetorical system is conceptual and cognitive and the reward system is based on recognition and reputation (p.440). Cozzens (1989) identifies the reward system as founded in normative citation theory and the rhetorical aspects as contributed by the tactical theories of citation. Then she re-defines the two systems in terms of their convergent aspects; these aspects cut across divisions as they are found in interpretations and operationalisations of attribution theories (2.6). The factors that are identified as influencing the likelihood of citations fall into two groups,

those that are attributes of how documents connect rhetorically to others in their document networks and those that are attribute of the relationships of the authors to their competitors (p.442)¹²

¹² However operationalisation and differentiation is still difficult on the basis of these two groups as can be seen in the findings of Baldi and Stewart [Section 2.6.1.1].

Cozzens (1981) also introduces a third system, the communication system (p.444). Though the communication system is left relatively underdeveloped in Cozzens' model, this inclusion allows recognition of the role of institutional mechanisms associated with citing practice and places the mechanisms of the communication practices of science in direct relation to citing practice.

Among the citation inflators and deflators, journal characteristics, language of publication, and other measures of audience size need to be analyzed as part of this system (p.444)

Though the communication system is no less concretely distinguishable than the other two systems, its introduction does assist in allowing Cozzens to discuss the possibility of addressing predictability in citation behaviour, through the 'patterns (habits, conventions) of incorporating the contents of documents into text that make for different levels and temporal profiles in citations' (p.442). These patterns again allow for the possibility of a citing theory that can be developed to support further citation analysis and citing theory development that includes statistical modelling, representing the possibility of attempts to overcome quantitative and qualitative divides within much of citation theory development (p.445).

Cozzens (1989) is using an interpretation of the tactical theory of citation that focuses on the function of documents to argue knowledge claims. If a knowledge claim is to be accepted, its argument must be persuasive, and referencing is one of the rhetorical resources authors marshal, both conceptual and honorific, to achieve acceptance of their work through support of their argument and resulting knowledge claim. Referencing is an aspect of the writing process of an author, which has central communicative functions within the reward system of science (p.445). The function of the reference is to incorporate 'the contents of documents into text' (p.442), thereby creating a meaningful connection of the knowledge claim understandable to its audience, it is the 'meaning (or meanings) citing authors attribute to the document that determine its citation pattern rather than the author intent' (p.443). Placing citing practice in the context of the writing process allows for a reconciliation of the 'context of justification' and 'context of discovery' polemics. Writing is, for example, argued to be simultaneously justification through communication (Bazerman, 1983) and discovery as a 'knowledge-constituting process' through text production (Galbraith, 1999).

Without reference to Cozzens' rhetoric-first model, Nicolaisen (2003) argues from a philosophical and documentalist perspective the necessity of regarding citation practice

as part of the scientific writing process within theories of citing. Citing is a 'social act' in Nicolaisen's model and is 'closely connected to the social conventions of writing traditions', which for an individual author are determined by Laudan's "research traditions",

in order to understand, explain, and predict the dynamics of citation networks we need to penetrate the social worlds of individual authors (p.18)

Nicolaisen stresses the social elements of the act of citing as he argues current theories that consider writing, do so only from a cognitive perspective (citing (Small, 1987)).¹³ When Cozzens' rhetoric-first model, and its foundational positions anchored in the tactical theories and the normative theory of citation, is also considered, it is apparent that the social aspects of the writing process are to some extent effectively supported in this way from within previous citation theory. The role of such collective groupings such as 'research traditions' are only properly supported through the convergent aspects of the two attribution theories.

Cozzens' rhetoric-first model of citations can be reconciled with approaches to the study of citation practice that involve scientific writing as a social, communicative and cognitive practice. Hyland (1999) develops such an approach; though he does not use Cozzens' rhetoric-first model of citation, there are foundational similarities between them. Citations in Hyland's rhetorical account 'display an allegiance to a particular community or orientation, create a rhetorical gap for his or her research, and establish a credible writer ethos' (p.342). 'Intertextuality' is a prominent feature in Hyland's account; citation contributes to the way a writer situates 'their research in a larger narrative' (p.342), and so includes consideration of the expected audience for the work. Academic writing is seen as a tension between 'originality and humility to the community, rhetorically accommodating laboratory activity to the discipline' (p.342).

citation helps to define a specific context of knowledge or problem in which the current work is a contribution...New work has to be embedded in a community-generated literature to demonstrate its relevance and importance and to accommodate readers' scanning patterns as they rapidly search for relevance and newness (citing (Berkenkotter & Huckin, 1995)) (Hyland, 1999, p.343). [citation practices] contribute to how writers choose to frame their studies for colleagues, relying on a sprinkling of citations to invoke a set of common

¹³ It is noted that in his later work Nicolaisen (Nicolaisen, 2004, p.89) does acknowledge that Small also considers the social in his formulation. However, contrary to Nicolaisen's interpretation Small does not appear to have done an about face in relation to focus on psychological aspects, rather his approach was foundationally embedded within Mertonian sociology of science which considers science originally as primarily social as well as cognitive; however this evidences that interpretations of normative theory of citations have been thought to emphasise the cognitive aspects (at times to the exclusion of social elements).

understandings through to an elaborate scaffold of supporting references (Hyland, 1999, p.362).

Hyland completes a rhetorical analysis of citation use, using broad disciplinary categories, and determines that citations reflect disciplinary differences and that the 'impact of citation choices clearly lies in their cognitive and social value to a community' (p.362).

Paul (2000) undertakes an analysis which utilises Cozzens' rhetorical model of citing, and so combines rhetorical (processes of selection and positioning) and reward (community acceptance) functions of citations. This study is a rhetorical study of the 'multidisciplinary' field of chaos. Paul uses quantitative content analysis methods to examine intertextual relationships between both citing and cited texts, rhetorical strategies used by the citing texts, and by using citation counts as measures of community acceptance. A document set of citing articles is developed from 13 articles produced by 2 groups within the Chaos field, and the examination is of how these articles are cited across 3 time spans between 1975 and 1994. The time spans were identified using measures of growth as chaos was represented in the journal 'Nature', the time spans are then aligned with three periods of development. Overall Paul finds that the rhetorical use of citations changes according to the developmental time spans identified and that articles are cited in an increasingly specialised way, which indicates to her evidence of 'an increasingly stable and specialized community' (p.213). Paul's findings support Cozzens' rhetoric-first synthesis as it contributes to the explanatory value of understanding of citing practices that combines aspects of both the normative and tactical theories of citation.

Cozzens' synthesis of the tactical and normative theories demonstrates an attempt to address the structural and agency divide that occurs in operationalisations and discussions of the earlier theories. The work is supported by empirical attempts to address these problems through convergent theoretical positions; however these empirical results are also singularly concerned with individual citing acts, so they do not in practice address the positioning of a communicating author in terms of their broader institutional environments, therefore do not place the scientific work of the individual within context.

2.8. ASSOCIATION THEORIES OF CITATION

The combined dual interpretations of the normative citation theory indicate that it is important to be able to allow for accounts of citation from the perspective of individual

authors citing practice as well as patterns and distributions produced from the aggregation of citations. However it has been argued that current formulations of the attribution theories of citation do not provide for this concurrently. This means that it is very difficult to produce theoretical accounts of individual acts of citing and individuals' citing practices as they are related to broader knowledge production systems across multiple levels of aggregation. The following functional theories introduce accounts of citation and citing that attempt to achieve this. They are all concerned with the centrality of communication and its associated role of representation in scientific work practices. The importance of recognising the dyadic relationship between citing-cited artefacts is also considered within these theories, though addressed in varying ways.

2.8.1. SYMBOLIC 'THEORY' OF CITATION

Following Garfield's (1977) "citation markers" and reflecting Merton's (1988) symbolic functions of references, Small (1978) introduces his account of 'cited documents as concept symbols'. This is partially in response to the growing number of studies that were focused on the 'why' of citation and citing. Most of these accounts are motivational in focus but also included categories representing the functions of citations. The crucial aspect re-introduced within Small's symbol theory is the centrality of 'content' in citing behaviour. Citations as symbols of concepts or methods, have the cognitive function of embedding the scientists-authors work in earlier literature, through referencing (p.328).

Citing is a symbol making process, where 'in citing a document an author is creating its meaning' (p. 328). Small aligns this process with language, labelling and writing in scientific papers (see also (Small, 1987)). Through citing a document an author 'establishes a link between the cited document and the language in the citing text' which is expected by the reader to have some rationale. This conceptualisation allows for a process whereby private ('nonce') symbols are transformed into public ('standard') symbols, through the cited symbol acquiring a 'standard or conventional interpretation (meaning)', it becomes a 'collective representation' (pp.338-339). Though this transformative process is not part of Small's empirical study in this paper, he does discuss the possibilities of recognising it. Viewing citations and citing in this way acknowledges the dual role of citing. Citations are both cognitive and socially determined, 'citors engage in a dialogue on the document's significance' and 'the process of acquiring a standard or conventional interpretation is crucial for the social determination of scientific ideas' (p.338).

For Small (1978) citations only become ‘standard symbols’ through aggregated use by many actors (p.329),

my suggestion is to graft a network epistemology onto the normative theory of science, borrowing methods of textual deconstruction from the constructivists, but without adopting their relativism. Citations then become part of the process of argumentation, justification and interdisciplinary bridging, with the end product being the consilience of scientific knowledge (Small, 1998, p.147).

This is empirically examined through his notion of ‘uniformity of usage’ or ‘consensus’, defined as ‘the percentage of citing contexts which share a particular view (the most prevalent) of the cited item (Small, 1978, pp.329-330). In ‘this first empirical step’ he finds reasonably high uniformity percentages (only one concept is less than 42% (p.335)) and interprets this finding, in terms of citation exemplars, ‘as evidence of an author’s compliance with the general paradigm’ (p.338). Using Small’s symbol theory of citing and Gilbert’s suggestion ‘that research networks tend to develop standard references’ Cozzens’ (1985) finds partial empirical support for meaning consensus development across different fields.

Budd (1999), finds evidence that supports Small’s theoretical symbolic account of citation practice through examining citations as epistemic links between knowledge claims in the scholarly communication process, and interprets the implications for this concept as the necessity of acknowledging ‘the interaction of author and reader’ (p.267). Meaning within a text is determined via interpretative processes for both reader and author, so the ‘author and reader are engaged in a dialectical relationship’ which can be seen as residing in a broader dialectical discourse’ (p.267). Citation practice involves epistemic and non-epistemic¹⁴ linkages, that citations are part of an author social and epistemic [without falsity] justification within scholarly communication, and that citations can be seen as ‘contextual extensions of the author’s argument of proposition’ (p.272-273).

Small’s symbolic account of citations allows concerns such as intertextuality to be considered in relation to the functions of citation within communication processes. Recognition of the significance of seeing citing-cited entities as *relationships* has been incorporated into further work. Baldi (1998) develops his analysis based on ‘citations as dyadic relationships’, finding that empirical distinctions between existent motivational citation theories are not clear, and elements of both are evident when seen as relationships between citing-cited documents. This has produced evidence that Small’s

¹⁴ Budd (Budd, 1999) defines non-epistemic primarily as including links that could not be verified through explicit evidence as epistemic within his analysis

symbolic theory of citation moves beyond divisionary interpretations of the attribution theories of citation.

Small's symbolic theory of citations is able to provide an account for the variation in findings related to relevance between citing-cited documents in empirical citation studies. Harter, Nisonger and Weng (1993) found that subject content and citing-cited relationships typically have only a small amount of similarity. This finding empirically argues against the use of citation indexes in identifying subject areas and citing-cited pairs having content based linkages. However, the Harter, Nisonger and Weng findings are countered by Peters, Braam and van Raan (1995) in their analysis of cognitive resemblance also operationalised through content keywords and citing-cited relationships. Peters, Braam and van Raan (1995) find that that publications with a citing relationship are content-related, and those that share a reference to a highly cited paper are also content-related to a significantly higher degree than documents that are not bibliographically coupled.

The conflicting findings of these studies by Harter *et al* and Peters *et al* are further complicated by the findings of Braam, Moed and van Raan (1991). When compared, co-citation clusters produce low recall in relation to subject content retrieval, and so co-citation clusters are considered not to be an effective tool for retrieval across a specialty as compared to subject based approaches (Braam et al., 1991). This result implies that the identification and definition of specialty boundaries when using co-citation methods is not congruent to identifying a content defined specialty area. The diversity of findings represented in these studies finds theoretical explanation in Small's symbolic theory of citation. Small's theory predicts a non-linear content relationship between citing-cited pairs of documents via the process of citing authors' interpretative role in establishing the epistemic link between the documents. The content relationship is associative and not necessarily encompassed completely by topical relationships, but is related to them interpretively through consensus in meaning. Budd (1999) is referring to these processes as epistemic and Small (1998, p.144) refines this as a 'network' or 'social epistemology' (citing (Schmitt, 1994)).¹⁵

¹⁵ Harter (1992) develops a psychological account of the relevance judgments made during citing. Relevance judgments made during citing are not restricted to topical relevance and so are not necessarily captured through subject analysis, but rather are related to the research process of the citing author, and their psychological state. Nicolaisen (2004) criticizes Harter's account of citation relevance based on the singularly mentalist approach he takes. However when Small's symbol account of citations is seen as confirmatory of Harter's psychological relevance, citing practice is simultaneously socially influenced, with the recognition that simultaneous content and semantic process are involved.

Implications of the author-reader relationship and the interpretive nature of citing practice are recognised by Small and supported by Harter's (1992) psychological relevance. Both conceptualisations relate to the necessity of a semantic understanding of the citing-cited entity relationship which can involve, but does not necessarily, direct and simple content (as in specific word) relationships. Accordingly Small (1978) notes that motivational aspects of citation theories, and too narrowly defined content focused interpretations of citing practice, are not sufficient to account for their functions within scholarly communication:

The relationship between the cited document and the concept it symbolizes...is "metaphoric". In the extreme, this means that there need not be any similarity between the document and the concept it stands for...the perceived content of the document is independent of the document itself...(Small, 1978, p.329)
The interesting question is not whether the cited work is "correct", or whether the citing author has made a "correct" interpretation of it, but rather whether the interpretation is in accord or at variance with the interpretations others have given it (Small, 1978, p.338).

Van der Veer Martens (2001) extends Small's 'theory' of citations as concept symbols by relating it to Star and Greisemer's (1989) "boundary objects". For Van der Veer Martens this conceptualisation of citations as boundary objects introduces recognition of the 'plastic' and 'robust' nature of citations as they function in 'local informational needs' and to 'maintain a common identity in several intersecting social worlds' (p.2). She further relates these diverse functions to the correspondence, coherence and consensus "theories of truth", finding support for these functions in examining scientific communication, United States legal and patenting system processes, at the micro, meso and macro levels. Focus in her later paper (Van der Veer Martens & Goodrum, 2006) shifts to using a combination of functional citation context analysis and in-depth interviewing of citing authors as well as relevant journal editors to examine the functional role of theories within knowledge production. Though Van der Veer Martens explicitly says she is not contributing to 'the extensive literature on citation behaviour' (p.334), functional heterogeneity in the use of references is apparent in her findings and these have relationships to functional use of theories. She finds evidence that reference functions within scientific literature correspond to aspects of theory diffusion and its use, according to applicability, constructivity, accessibility, connectivity and generativity, as supported by the other methods of analysis.

Small (2004a) further develops his symbolic theory of citation through identifying its placement in the 'common ground' derived from both the normative and tactical theories of citation,

if the norm of citation involves a symbolic payment of intellectual debt, it is, at the same times, an ascription of meaning to the cited text and a construction of its meaning (Small, 2004a, p.76)

Small (2004a) then develops a possible methodological operationalisation to support this postulated balance through operationalising the normative and tactical theories of citation through how much citations represent convergence and divergence from consensus meanings.

When the constructed meaning is coincident with the author's original message as well as common usage, we may say that there is a strong normative compliance...When the constructed and original messages diverge, we have what is for a normative sociologist deviant practice on the part of the citing author, but perhaps normal behavior for a constructivist (p.76).

This is certainly an interpreted use of the attribution theories of citation, but one that does recognise the possible theoretical and empirical convergence associated with them. At the same time the question of ascribing intent to behavioural analyses remains quite appropriately, specifically and deliberately unasked and unanswered in this operationalisation. Empirical analysis using this operationalisation is left unreported in this presentation and currently remains empirically untested.

An account that both theoretically and empirically extends Small's symbolic theory is presented in (Small, 2004b). This paper develops the symbolic theory through introduction of its possible applicability to authors' views on their own citation success. Authors' 'opinions' are accessed using the 'Citation classic commentaries' archive available through Essential Science Indicators (ESI) Special Topics produced by Thompson Scientific. In the course of contributing to these commentaries the authors also comment on their understanding of the significance of their highly cited work. Small recognises the potential difficulties associated with analysis based on using authors' opinion as presented in this public forum. However, the functional categories found through content analysis of these commentaries vary from, but are well supported by, other functional analyses using different methodological combinations that are somewhat less 'public' or with less implied self-interested motivations (for example (Amsterdamska & Leydesdorff, 1989; Leydesdorff & Amsterdamska, 1990)). The non-exclusive categories Small identifies in his sample of commentaries include: Interest, Novelty, Utility and Significance. These are generated from a content analysis of responses to open requests for commentary rather than specific questioning of authors in these areas. Interest is the category that forms the central position in relation to utility, with novelty and significance forming a triangle of relationships through the interest category. What is demonstrated in this study is the necessity to

consider the social and cognitive significance of placement of research within frameworks in their relevant fields. Determination of citation characteristics requires papers to be established as they are embedded within their 'fields' or 'specialties', as key functional relationships to the field, such as importance and significance, are determined through these structural levels.

Small's deliberate focus is on accumulation of meaning represented in the convergence and divergence associated with aggregated use of concept symbols (cited documents) (Small, 1978, 1986, 1987, 2004a, 2004b). It has been argued that semiotic frameworks and their associated concerns, such as intertextuality (Budd, 1999; Hicks & Potter, 1991), symbol-sign relationships (Cronin, 2000), and dyadic relationships (Baldi, 1998), require that the author (individual researcher) be seen as an interacting and participating agent in the citing-citation process (Hicks & Potter, 1991).

2.8.2. SELF-ORGANISING COMMUNICATION SYSTEMS 'THEORY' OF CITATION

Leydesdorff's (1987) rationale for his functional theories of citation argues against both Cozzen's rhetorical theory and other 'sociological' approaches based on the 'citation practices of scientists', as they do not adequately account for the 'cognitive perspective' (p.306). Focusing too singularly on citing actors leads to the exclusion of their cognitive role within the 'self-referencing social systems of science', removing them from the substantive reasons for their behaviour. Leydesdorff places citing actors within a broader system of accepted scientific knowledge, which is both social and cognitive:

'citations' as acts of scientists, are both forms of behaviour and at the cognitive level carriers of cognitive linkages among different knowledge claims, or among knowledge claims and accepted (codified) scientific knowledge. (Leydesdorff, 1987, p.306)

Social order and structuring has a complex relationship to its constituting actors within Leydesdorff's theories; order emerges 'from networks of communication relatively independent of carrying authors' (Leydesdorff, 1998, p.13). However the systems are expressed through 'densities of selections' that result from, though are not completely accounted for by, the accumulative actions of these actors (p.20). Levels of aggregation in the communication system are not entirely commensurable, and as such theoretical perspectives and analysis based on the varying levels need to be developed with this in mind. Leydesdorff maintains incommensurability amongst levels of aggregation is a feature of self-organising systems, of which science is an instance, and that this does not preclude the possibility of development of citation theories that account for citation practice and support validation of citation analysis:

[citing is a] dynamic operation that allows for reduction of complexity in various contexts at the same time. The dynamic perspective of selections operating upon selections in other networks accounts for the character of citations as statistical (uncertain) indicators, for their specificity, and for their multi-contextuality. (Leydesdorff, 1998, p.7).

Leydesdorff's later theoretical positioning derives from attempting to account for the diversity in citation functions found in earlier empirical work with Amsterdamska on the epistemological functions of citations in knowledge claims (Amsterdamska & Leydesdorff, 1989; Leydesdorff & Amsterdamska, 1990). Though Leydesdorff and Amsterdamska begin their analysis with selecting a set of four cited articles, they are analysing the 'use' of those citations in articles that cite them. They are in effect analysing citing behaviour and the role 'cited claims play in the arguments which are being constructed in the citing articles' through citation context analysis (Amsterdamska & Leydesdorff, 1989, p.452). Functional relationships between the citing and cited articles are emphasised.

Four functions of cites are used as a basis for analysis: transformation of the knowledge claim, warrant (support) of the construction of the knowledge claim, agenda building or to legitimate the importance of the research interest, and contextualisation of the article within the knowledge structure of the field (Amsterdamska & Leydesdorff, 1989, p.455). Variations were found across these categories in the way cited articles were used in their citing articles. Amsterdamska and Leydesdorff interpret from these findings that cites 'constitute a heterogeneous category at this level of analysis' (p.460), and differ significantly in the way they attempt to integrate claims into their field of knowledge (p.461) and so will violate assumptions of structural equivalence within network analyses based upon their use (p.468). The conclusions in this earlier presentation of this work are further supported in the findings of heterogeneity in the perceptions of use of cited articles by citing authors as well as their evaluations of the research group that produced the cited articles examined, according to the categories of quality, relevance, influence and originality (p.317). However their results indicate that in questionnaire responses authors do not discriminate in their answers in ways expected from their detailed construction of functional categories (Leydesdorff & Amsterdamska, 1990, p.320).

Leydesdorff and Wouters (1999) begin to develop a theory that attempts to address the duality inherent in a semiotic system of citations as well as account for the relationships between the subtext (reference) system and super-text (scientific literature) system

within scientific communication.¹⁶ In this account references are the selective functions in evolving systems of scientific communication through the process of recursive selections (pp.170-171). Selections form the coupling between the subtext and super-text systems and so can be 'used to increase the certainty of the outcome at the next level, in other words to improve the expected information content of these literatures' (p.172). The recursivity of selections produces the epistemological role of referencing in scientific literature (p.172). There is an inherent friction in the coupling of the two systems of referencing and citation; however selections serve to indicate the mutual interactive shaping of social and cognitive dimensions in the co-evolutionary scientific communication process, allowing examination of that process across and within the two systems (p.174).

At the centre of Leydesdorff's dual systems of references and citations is the scientific communication system. The socio-cognitive elements are brought together within an overarching system of scientific representations and their communicative interactions and consequences. In order for a more complete articulation of the inherent duality of reference-citation systems in citation theory, Leydesdorff postulates a reflexive meta-theoretical approach (Leydesdorff, 1998). The approach is conceptualised as a dually layered 'multidimensional network' in which 'citations are the result of the interaction between networks of authors and between networks of their communications' (p.9).

In recognition of the multidimensional nature of the networks involved, Leydesdorff postulates that multiple theories of citations are required to account for it. The multiple theories are necessary to account theoretically for the empirical imperative of specification from among multiple 'windows of observation' (Leydesdorff, 1998, p.20). The theoretical possibility of a 'multitude' of 'apparently incompatible citation theories' is not considered to be a 'problem that must be resolved' in Leydesdorff's conceptualisation; this is rather seen as 'an opportunity to forge more complex relationships within the system of scientific communication' (Leydesdorff & Wouters, 1999, p.179). The central features of Leydesdorff's theories of citation and postulated meta-theory are most explicitly articulated in his series of papers (Leydesdorff, 1993, 1998; Leydesdorff & van den Besselaar, 1997). Figure 2.1 is a visualisation developed

¹⁶ Wouters' (1999b) emphasizes the fundamental significance in differentiating between citations and references. This differentiation indicates that citing behaviour 'cannot, contrary to received wisdom in scientometrics and science studies, be regarded as sufficient to explain the role and function of the citation' that it is 'the symbolic process at work in citation indexing needs to be analyzed' (p.12). Wouter's (Wouters, 1997) argues that the transformation from citing sign to citation sign is so fundamental that the two exist in different systems (pp.496-497).

from these papers that attempts to capture the interactivity between systems crucial to this theoretical position.¹⁷

¹⁷ Leydesdorff's work contemporaneous to the visualized theories (Figure 2.1) extends the foundations of his theoretical position, providing more detail in self-organization, social order and the complex dynamics of scientific communication (Leydesdorff, 1995). His later work develops this foundational work further, detailing the significance of the interactive nature between scientific practice, cognitive and epistemological codification and the role of communication in scientific actor-networks (Leydesdorff, 2006), also providing a simulation of these processes in (Leydesdorff, 2005).

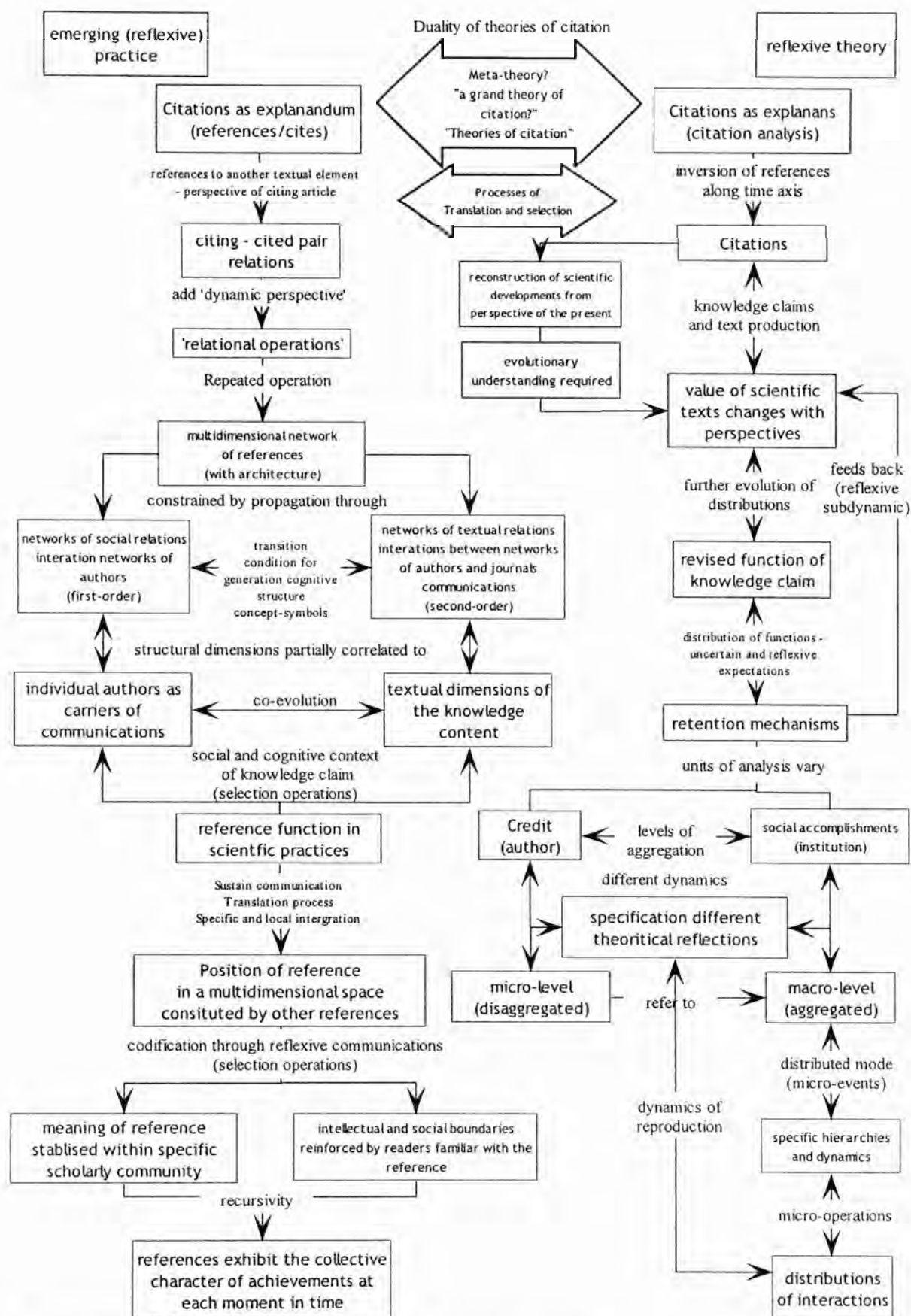


Figure 2.1 Visualisation of Leydesdorff's "Theories of citation"

van Raan (1998) critiques the possibility of a 'citation theory'. However the particular definition he is using in this paper is that of 'citation theory' as motivational citing

theory. In his view the basis of the problem of developing a theory that supports citation analysis from motivational and heterogeneous characteristics of citation practice is that citation analysis is concerned with ‘an ensemble of many citers’, not with individual citing acts (p.136). There is no straight-forward relationship between ‘the global characteristics’ and the micro level act of citing, as global characteristics ‘are typically independent of the micro-characteristics of the constituting elements’ (p.137). It is the ‘distribution-function of these characteristics’ that are ‘the make-up of that part of the world which is relevant to bibliometric analysis’ (p.136). Leydesdorff’s work is introduced by van Raan as an account that holds potential for recognising and overcoming this central difficulty. Through an analogy between Leydesdorff’s model and chemistry, van Raan interprets citations as the ‘binding properties’ of the scientific literature’ and the system is self-organising on the basis of the formal linkages citations provide in the scientific communication system (p.137). It is this functional role of references-citations that allow distributional placement of citation practice within ‘a larger whole’ (p.137).

Despite van Raan’s endorsement, Leydesdorff’s theory only postulates a possible connection between the dual citations systems via a ‘meta-theoretical’ approach; empirical accounting for it still remains at issue. Therefore the role of the individual within knowledge production remains to be determined through simultaneous accounting for behavioural and epistemic micro-theories within Leydesdorff’s meta-theory and theories.

2.9. THE SITUATED KNOWLEDGE PRODUCING INDIVIDUAL AUTHOR

As noted at the beginning of this chapter, it is acknowledged in the field of citation studies that a comprehensive theoretical account of citation practice is yet to be formulated. This review has examined the literature on citation theories and has argued that the individual citer as situated communicating knowledge producer remains to be accounted for within theories of citation, and that this a necessary addition for citation theory in order to adequately consider the work practices in which citation is embedded. Attribution theories of citation address the individual in very different ways and do not singularly support interpretation through citation analysis of their role in terms of scientific work practices in knowledge production. Associative theories of citation may be able to account for the individual theoretically, but a practical means of assessment has not been developed.

The next chapter discusses scholarly identities and their relationship to citation identities, and preliminary identification of requirements to account for scientific work practices and the positioning of an individual within knowledge production communities.

3. COMMUNICATING IDENTITY AS A SCIENTIFIC WORK PRACTICE - SOCIOCOGNITIVE REPRESENTATION OF THE INDIVIDUAL KNOWLEDGE PRODUCER

The previous chapter highlighted the lack of sufficient support in citation theories for developing further insight into the role of the individual concurrent with their structurally dependant scientific work practices. The examination of citation theories and how they currently relate to individuals; however, did indicate a number of important considerations for what such a theory should be able to address.

Consideration of the duality of the semiotic aspects of citations should be addressed through noting the significance of the relationship between citing-cited pairs. Various levels of aggregation are not easily reconciled within current theory, and it is demonstrated that this is significant in allowing for adequate description of the role of an individual within their knowledge production environments. The associative theories indicated the importance of considering the individual as part of the communication system, and that this was an integrated process within knowledge production. The normative theory of citation emphasised the importance of the institutional structures that are fundamental to the communication system. The tactical theories demonstrate that agency has a central role in producing the representational outcomes of scientific work practices and that citation practice is part of this larger communication process.

Related to an individual researcher's role as a communicator there was also emphasis on the writing process within scientific work practices. Citing has a significant role within the scholarly writing and knowledge claim processes, though understanding of this role is not well addressed in theories that allow for distribution and citation pattern justification. A citation theory that adequately addresses the role of the individual and their scientific work practices in knowledge production ideally will be able to account for both the individual's production of knowledge and their placement within broader scientific communities.

This chapter reviews the literature in relation to the conceptual possibilities raised in White's citation identities and author publication profiling, Lievrouw's scholarly identities, and the conditions of the relationships between an individual researcher and their knowledge production units which determine their scientific work context, content and practices. Further, how these theories might begin to address some of the directions indicated by citation theory and studies is examined.

3.1. CITATION IDENTITIES AND STYLISTICS

White's work on citation identities, author stylistics and CAMEOs provides the framework necessary to begin examining the researcher/scientist as producer of stabilised representations, and allows for these representations to be placed within their knowledge production networks.

White's (2000) 'Toward ego-centered citation analysis' introduces his conceptualisation of citation identities, and connects this technique with a broader profile of an author's contribution. Through a detailed analysis of Eugene Garfield, the person to whom the whole volume is dedicated, White introduces four potential 'ego-alter' variables, extending concepts from ego-centric analysis (a subset within social network analysis) into citation studies. Of these four 'ego-alter' variables this study will focus predominantly on the development of citation identities; the other three ego-centered variables include co-authors (as an operationalisation of collaborators), citation image (co-cited authors or authors jointly cited with ego) and citation image makers (citors or authors who cite ego).

White (2001b) further establishes the concept of citation identities and more specifically articulates it, combining the conceptualisation of citation identities with broader socio-cognitive analyses of individual authors' careers. White's examination introduces analysis of a variety of components contributing to the understanding of individual authors' publication careers. This study established a number of variables seen to be contributing to the production of an individual author's 'cognitive profile', including co-authorship, citation profiles and citing profiles. The profiles developed were then further analysed according to biographical and historical information on each of the authors. Through this process a number of features of the publication careers of these authors was elucidated, culminating in the development of an initial typology of identifiable characteristics of the authors' careers. Examples of these characteristics are 'interdisciplinarity', 'divided careers', subject relevance' and 'controversy'.

Further extending the notion of bibliometric author profiling are White's (2001a) CAMEOs (Characterizations Automatically Made and Edited Online). Here White also considers the use of subject descriptors, natural language keywords, journal categorisations, journal names and publication years ranked by productivity, increasing the potential number of access points into an individual author's representational profile. This allows profiling techniques to be used in other databases besides the

citation indexes, potentially producing a more varied description of the author's publications than those possible with using the citation indexes alone. In effect this methodology introduces a non-intrusive comprehensive suite of data-mining possibilities with which to examine and analyse authors' careers as represented through multifaceted publication profiling.

Following White, Cronin and Shaw (2002) use selected elements of CAMEOs and 'thick description' of three information scientists to develop rich 'biographical sketches of authors', drawing together authors intellectual, social and institutional affiliations articulated via an author's individual writing and citation style, to form what they refer to as a 'watermark' of authors' 'scholarly output'. Due to the nature of this study the sample size is also small and is formed by a cohort of Cronin's and Shaw's peers. Though offering a deep picture of the authors under study that would otherwise be extremely difficult to obtain, this paper does not address implications for citation theory.

White's (2001b; 2000) primarily conceptual and methodological papers, firstly examine only information scientists, which is only one model of disciplinary scientific work practice, and secondly the sample sizes in both studies are limited to one in the first instance and eight in the second. So these initial studies are limited in terms of size of sample and in disciplinary representational coverage, a point White himself readily recognises and contextualises in terms of their illustrative and primarily conceptual purpose as presented in these papers.

A central feature of White's (2001b) analysis of citation identities and subsequent stylistics is the incorporation of his understanding of the field of information science, both as a participant and as a researcher. White's earlier work with McCain (1997; 1998) in the mapping of information studies is drawn on as well as his personal knowledge of the researchers he is investigating. This extensive knowledge of the area is useful in developing the potential of citation identities to describe individual researchers in this exploratory study, and indicates the utility to be derived from drawing on a combination of qualitative and quantitative approaches in this work.

Analysis of citation identities themselves are further elaborated through the juxtaposition of an author's frequency of citing against their use of citees (or those authors that they cite). When these two features are contrasted against one another a unique pattern of use and reuse of citees becomes apparent within an author's citing distribution (White, 2001b). From analysis of these patterns White begins to examine

and raise the interesting prospect of an author's citation style. The concept of an 'author's style' is an extension into informetric profiling of authors derived from the work of Herdan (1960), the quantitative linguist who utilises authors' use and reuse of words to begin to identify or differentiate between them. White extends this concept into an author's pattern of usage of citees versus citation, and thus moves from Herdan's Type-Token ratio into his Citee-citation ratio.¹⁸

As it is developed within White's suite of techniques, the Citee-citation ratio becomes a significant measure, allowing both distributional and content features of an author's publication outcomes to be examined as they are inter-related within an author's publication career. Concurrently the author's placement within the broader institutional structures they are embedded in can be addressed. This allows White to develop a distribution based analysis of citing behaviour across and within authors' careers, representing the authors' usage of prior literature, to elicit a picture of an individual author's 'cognitive stylistics'. Importantly this is comparable between authors. These 'cognitive stylistics' provide a mechanism with which to tie distributions of individual behaviours back into citation theory. In effect this technique introduces a simple measure of concentration, though it is not articulated as such within White's paper.

The conceptualisation of an author's citation style and behaviour as concentration of use of citees and citations enables White to identify patterns of 'core-scatter' within authors' citation oeuvres. The 'core-scatter' patterning that White (2001b, p. 95) establishes from a combination of use and reuse analysis and the citee-citation ratio, is used to identify three citation styles: 'scientific paper' ('relatively few authors; much recitation'), 'bibliographic essay' ('relatively many authors; little recitation') and 'literature review' ('relatively many authors; much recitation'). It is made evident in the course of the analysis and interpretation presented that individualised author citation profiles are formed from the interplay between features of diversity and concentration of citation patterns within the representational outcomes produced by the researchers. The diversity-concentration patterns evident are the outcomes of the researchers' use of prior literature forming a unique accumulative representational profile, produced as a result of the researchers communicative and scientific work practices.

¹⁸ White's use of Citee-citation ratio is in fact an inversion of Herdan's Type-Token ratio, thus equating with an average of citation to citees. This study reverts to the original formulation by Herdan, thus is calculated citee-citation. Rationale and usage of the Citee-citation ratio within this study is discussed in Chapter 4.

Citation identities are an important component of author characterisations in White's (2001b) conceptualisation and allow the author to be placed within their intellectual environment: 'they are significant in studying an author's intellectual history and whatever it portends for the author's field' (p87). Further, the author characterisations produced in citation identity analysis are predicted to be 'as distinctive as fingerprints' (p.88), therefore will uniquely identify the sociocognitive representational histories accumulated over authors' *oeuvres*. Citation identities are behavioural as opposed to motivational in the sense that 'it is doubtful that authors premeditate, or even notice, the citation patterns that form in their *oeuvres* over the long term'; they are the 'unplanned by-product' of many individual citation selection decisions by a particular author (p.88).

Neither the normative nor the tactical theories of citation are addressed in detail in the initial formulation of citation identities or stylistics (White, 2001b). However White recognises the empirical diversity found in citer motivation schemes, such as those developed by Cronin, Brooks and Liu, and develops a position that attempts to accommodate this diversity and address his citation identity findings. He frames this through the concept of 'perceived relevance'; 'the most important citer motivation is to *project one's writing* – particularly, to project the coherence of the *oeuvre* by binding earlier work to later within specific contexts' (p.102). 'Perceived relevance' is the underlying citing motivation that brings together the individual researcher as author, reader, citer and communicator (p.103). This interpretation of 'perceived relevance' is in many ways closely related to the associative theories of citation, in that it requires incorporation from cited sources into the citer's new work through developing its relationship to the work of the field; the perceived aspect considers interpretation a significant part of this process. The interpretation seems to suggest a behavioural rather than a motivational basis of explanation.

3.1.1. CITATION IDENTITIES AND THE NORMATIVE CITATION THEORY – WHITE'S ANALYSIS

White (2004b) extends examination of citation identities into their potential to interrogate existing citation theory, in an analysis of individual author's use of citees according to their 'reputation'. Citation identities are here utilised to empirically examine the attribution citation theories: the Mertonian normative theory, the tactical theory (constructivist theory), and critiques of the normative theory by authors such as the MacRoberts (as befitting an article within *Scientometrics*' special issue on Merton). The use of citation identities here is extended and refers to the reputational scales

derived from citation identities rather than to the identities themselves as previously defined in earlier work (cf. (White, 2001b)).

In White's (2004b) 'Sokal' study, citation identities are used to identify an author's citees and then the citations to those citees are collected across the three Institute for Scientific Information (ISI) citation databases: Social Scisearch, Arts & Humanities Search and Scisearch (1990-present).¹⁹ This procedure produces a career distribution of an author's use of citees according to how much the citee has been cited across 'science and scholarship', defining the author's use of 'reputationally' scaled citees. The citation identities for twenty-nine authors from a variety of disciplines and career stages are created; the authors include Alan Sokal for whom multiple identities are developed. From this examination White finds that the authors reference their citees according to a lognormal distribution, and 'a relatively equitable distribution of citations along the scale' (p.98), therefore the distribution is neither skewed to those at the low or high end of the reputational scale as determined by citations to the citees work across 'science and scholarship'. Those that are most frequently cited within these authors' citation identities are what White refers to as midrange citees (White, 2004b, p. 105).

White interprets his findings as consistent with Merton's normative theory of citing (White, 2004b). Central to this interpretation of findings is White's definitions of Merton's normative citation theory and the tactical theory of citation. In order to operationalise a determining distinction between the attribution theories, and recognising the potential difficulties involved, White further defines a fundamental distinction between the positions as involving 'dark persuasion' (tactical theory) versus 'light persuasion' (normative theory). These both imply use of authority to persuade but with different intent. Dark persuasion involves the manipulative use of excessive high reputation citees, which is translated into an operationalisation of 'name dropping'. Thus according to White's definition, inappropriate use of highly cited citees across science will provide evidence for dark persuasion, whereas use of citees of varying reputations within a citation identity will lend support to the norm of universalism and the normative theory of citation. White finds that there is not a disproportionate citing of 'world famous' as opposed to 'obscure' citees within these twenty-eight individual authors' citing careers, as would be expected within White's framing of Gilbert's persuasion hypothesis, or the micromotivationists in general; therefore his conclusion is that citing practice reflects a 'behavioural norm' (White, 2004b, p.109).

¹⁹ Dialog is the platform used by White to collect his data for this study, and he notes the attendant limitations such as retrieving only first citees of authors citing identity sets.

The difficulties raised in this account of the findings are consistent with those found within many studies that attempt to develop successful operationalisations of distinctions between the attribution theories (see Baldi, 1998; Moed & Garfield, 2004, section 2.6). The primary difficulty is the ascription of motivational interpretations on essentially behavioural studies. Successful persuasion according to both attribution theories and their accounts of authority amounts to similar behavioural citation patterns (Small, 1998; Zuckerman, 1987, 1988). The clearest demonstration of the problem comes from White's analysis of Sokal's citation identity and his hoax paper. Sokal's intention in this paper is not to contribute to the field in question but to make a point external to that. However in order to be successful he displays all the correct signage and content that allows for successful publication in a peer-reviewed journal. His intent was not evident to the reviewers of the paper nor is it demonstrated in the patterns emerging from his citation identity. In fact White makes the point that the lack of deviance evident in his citation behaviour for that paper contributes to the hoax's 'success' (p.113), therefore the question of the ability of citation behaviour studies to capture author intentionality remains, and is supported by the lack of distinction evident in Sokal's citation identities.

White's (2004b) normal citing reputational distributions can be seen more fruitfully as a reflection of the overall environment of the scientific community. Citations to scholars who both produce more work or that have been prominent within an area are most likely the ones who are going to be cited from both normative and persuasive theoretical explanations. The issue that remains is not how many world famous scientists are used within individual citing distributions, but if they are connecting the work of the author into an appropriate body of knowledge, as judged by their peers. Consequently if most scientists work within smaller communities of knowledge, (Chubin & Connolly, 1982; Chubin, 1976; Gläser, 2001b) it is their fellow scientists working on similar problems that should be cited in their work, i.e. those that are 'authoritative' within their scientific milieu. From an interpretative basis of 'perceived relevance' (White, 2001b), this will culminate in what is persuasive in both dark and light scenarios; the dark and light become in this way empirically indistinguishable. Therefore it is expected that scientists that are at either extremity of the obscure and world famous range would be represented less than scientists that may not have overall large or miniscule citation counts across all the sciences, but rather those that are well recognised within their own specialties/communities.

3.1.2. CONTENT AND THE SOCIOCOGNITIVE DETERMINANTS OF CITATION IDENTITIES

White (2004a; 2004) extends his account of citation identities through their empirical utility as demonstrated when used in combination with other methodologies, such as sociometric analysis (White et al., 2004) and discourse analysis (White, 2004a). They both examine more broadly the sociocognitive context in which citation identities and citations themselves are formed. Their findings indicate the centrality of content within citing practices when these practices are related, using these different methodologies, to their context in scientific work practices.

Sociometric and intercitation network analyses are combined in a study of 'Globenet' researchers (White et al., 2004). 'Globenet' is an interdisciplinary research group specifically formed with the intention of bringing together researchers from 7 disciplinary groups to 'promote interdisciplinary research on human development' (p.112). The sociometric analysis includes examination of the communication relationships between researchers, such as email and personal communication, self-identification of relationships between researchers, categorised according to friendship, acquaintance and colleague, and self-identification of level of relationship according to categorisations of collaborator, reading of each others work, discussing research and no affiliation. The intercitation analysis is developed from initial citation identities of the 'Globenet' researchers which are then used to develop intercitation matrices for the range of possible citing pairs between Globenet members. A third analysis is also introduced that provides an alternate identification of 'intellectual ties' through cocitation counts; this reflects how authors external to the Globenet network co-cite and therefore relate them to one another.²⁰

The first question White (2004) applies to this data is whether there is a primacy demonstrated between social relations or cognitive relations across the sociometric and intercitation networks. Two hypotheses are identified. Firstly, the social network hypothesis states

that members of an organized group should intercite considerably just because they know each other and that their intercitation should be strongly associated with the closeness of their various social and communication ties (White et al., 2004, p.112)

²⁰ The cocitation counts citation identities used to derive the intercitation matrices are taken from both the Science Citation Indexes and the Social Sciences Citation Index (Dialog files 7, 34, 434). The intercitation matrices are divided into four time periods (1972-1989, 1989-1992, 1993-1996, 1997-2000) and book data (1999) is necessarily manually collated. The co-citation counts are synchronous.

whereas the intellectual network hypothesis predicts

that intercitation is at most weakly associated with social ties and more strongly associated with a common discipline and shared subject matter

assuming

that social ties do not drive citation, they merely affect knowledge of items to be cited (White et al., 2004, p.112).

It is recognised by White that these analytical definitions are not entirely empirically distinct so a third hypothesis is introduced: the sociocognitive network hypothesis, where mixed ties are predicted and the relationships between researcher pairs involve both social and cognitive aspects and are primarily content-laden (p.112).

The Globenet study operationalises social and communication ties using the sociometric variables, perceived closeness and type of interactions between members. Intellectual ties are a combination of sociometric variables of self-selected disciplinary affiliation, reading of work relationships, and members' co-citation counts. Sociocognitive ties are operationalised using role variables, self-identification as a collaborator, co-authorship and editor or contributor to the book.

The findings of the Globenet study demonstrate that social and communication ties are significantly related to intellectual and sociocognitive ties. 'Being a collaborator, and, to a lesser extent, being a friend or reading the other person's work, that correlates significantly with both articles and book intercitation' (White et al., 2004, p.119). When these relationships are controlled on the basis of disciplinary affiliation, an increase through social ties developed in Globenet is not demonstrated for those researchers already known to each other in the same discipline, whereas for researchers across disciplinary affiliations social and communicative connections developed in Globenet did increase significantly. The Globenet study also finds that content as identified by cocited relationships and then compared to intercitation patterns, provides a definitional link between the social and cognitive activities of the 'Globenet' authors examined. Cocitation indicates 'broad commonality of subject matter, and authors tend to intercite in accordance with 'common subject matter'. This finding concurrently demonstrates the difficulties inherent in trying to differentiate the role of the social versus the cognitive factors in knowledge production, in terms of citation theory. Content, associative relevance (and attempts to accommodate the community and individual interpretation interplay) and its related sociocognitive ties, forms an important basis for such theory development.

The 'Globenet' study does not itself examine citation identities specifically, but rather extends the potential utility of examining an author's identity into a network environment through the concept of intercitation. Intercitation analysis focuses on the 'dyads' formed from the relationships between authors' citation identities. This is essentially a reference network analysis and a co-citation analysis which are then further analysed through a social network analysis from social relationship (via interview) data. It is the comparative nature of the resulting analysis that leads White (2004) to the findings that co-citation is 'the only significant predictor of intercitation for the articles data' (p.123); level of communication is reflected in citation, and both intercitation and co-citation have both social and intellectual components. Though the results and discussion of them strongly suggest that both social and intellectual ties are dominated by content relationships, White concludes that this is demonstrating 'the primacy of intellectual ties' (p.124).

White recognises the significance and complexity of content driven relationships in his study and discussion of the interactions within the fields of citation and discourse analysis (2004a). Beyond his discussion of motivational positions within citation theory, White e-establishes the necessity of understanding citation in terms of 'perceived relevance' (p.107).²¹ This global framework for developing an understanding of the motivational and behavioural elements of citing practice incorporates the imperative of understanding the multifaceted interactions inherent in content relationships between documents themselves and the environments in which they are produced.

The notion of 'perceived relevance' is not in itself without problems or complexity, but it does allow for the introduction and consideration of approaches that encourage citation to be considered from a number of appropriate standpoints, including their functional role, combining content and communicative elements within scientific work practice. Successful amalgamation of these factors concurrently and necessarily implies content-laden communication, as it coexists within the socio-cognitive mechanisms underlying knowledge production processes. Provision of an answer to the central questions of 'what persuades?' (p.109) requires 'regarding citation as a communicative process with syntactic, semantic and pragmatic variables' (p.112).²²

²¹ this is a re-statement of his position from (White, 2001b).

²² This paper (White, 2004a) in effect succinctly articulates the empirically and theoretically driven need for an associative, functional (or structural-functional) development of citation theory, (referring to the role of functional understanding of citation potentially derivable from Swales' CARS (Create a Research Space) model (p.106)).

To further develop the potential afforded by the empirical analyses made available through White's conceptualisation of citation identities, the communication practices of individual researchers and their development of representational identities is examined using Lievrouw's conceptualisation of 'scholarly identities'.

3.2. SCHOLARLY IDENTITIES AND RESEARCH NARRATIVES

For a number of years now Lievrouw has been developing a theoretical framework that identifies a central aspect of the scientific work of individual scientists as involving construction of research identities, which Lievrouw refers to as scholarly identities, through the mechanism of research narratives (see (Lievrouw, 1989, 1990, 1992, 1996)).²³ Her long term objective is to develop a systematic framework,

that describes the elements that comprise the narrative, the ways individual researchers may employ these elements to form and present their professional identities, and how narratives ultimately relate to the larger context of scholarship (Lievrouw, 1996, p.217).

Research narratives are the focal point from which Lievrouw's position on research identities depends, and are developed more comprehensively than the relationships between these mechanisms and the implications for research identities.

Lievrouw (1996) defines research narratives as

the scholar's own "story" of his or her professional life and work. It is a consciously constructed, strategic account that allows for the researcher to communicate the value and necessity of his/her work to peers and other important audiences (Lievrouw, 1996, p.216)

going further to link research narratives and research identities, a research narrative

is therefore a primary means of continuously shaping and communicating the researcher's identity, an important tool in what Mishler (1992) calls *identity formation*. [Lievrouw's italics] (Lievrouw, 1996, p.216)

Research narratives are the accounts and personal stories researchers develop for justifying and communicating 'their ideas, actions and strategies in the course of pursuing their particular scholarly interests' (Lievrouw, 1996, p.216). These accounts are not restricted to communication with research colleagues but also include friends, family, and acquaintances. Though there are differences in presentation and function across these audiences, the overall presentation of the researcher, their work and their

²³ Lievrouw's (1996) conceptualisation is underpinned explicitly by a constructivist perspective; later sections in this review introduce further conceptualisations that attempt to recognize the structural aspects seen as necessary to develop her account to accommodate for intellectual positioning. [see Section 3.4]

placement within the field is a central feature across these accounts. During interpersonal interactions research work and episodes are contextualised by the researcher ‘in terms of [their] whole career (or at least major interests, pursuits, or professional environment)’ (Lievrouw, 1996, p.217).

Lievrouw (1996, p.218) differentiates her conceptualisation of research narratives from Chubin’s “memoirs” (Lievrouw cites (Chubin, 1981)) and Woolgar’s “discovery accounts” (Lievrouw cites (Woolgar, 1976)). Unlike these forms of research accounting, research narratives are not restricted to prominent or influential researchers and are concerned more broadly with evolving accounts of the individual’s research as communicated throughout a research career than with singular focus on discoveries.

Research narratives are ‘considered as a specific type of scholarly communication behaviour’ (Lievrouw, 1996, p.217), and are communicated within all aspects of an individual researcher’s social representation of themselves and their relationships with the perceptions of others, as well as the individual’s anticipation of the nature of that relationship. Throughout her presentation of research narratives, Lievrouw focuses on the communication of research narratives in their ‘informal mode’ (p.217). This allows emphasis to be placed on the interactive and dynamic aspects of narrative construction. However, research narratives are not restricted to the informal accounts presented by researchers; they also involve more formal communications of their work through publication. A distinction between the various communication activities, self-presentations and narrative construction of researchers is not made, except to highlight that formal publications (such as journal articles) involve the added considerations of conventions particular to the form of publication. Lievrouw discusses a number of these conventions including the imperative to present research findings in an “objective” form, and the scientific ‘writing dictate that the author present ideas in a way that plays down his/her personal feelings about or stake in them’ (Lievrouw, 1996, p.220). Communication via formal publications is included in the conceptualisation of research narratives; however this aspect of communication of research narrative needs to be incorporated within a broader context that also considers other communicative activities that are less likely to be so directly influenced via research writing and publication convention.

Lievrouw claims research narratives are the mechanism through which researchers communicate and establish their research identity. An individual’s research identity is the dynamic result of how researchers ‘understand their work and its function relative

to what others do' (Lievrouw, 1996, p.216). Lievrouw (1996, p.222) uses Gergen's account of self to distinguish her conceptualisation of research identity from non-social or purely mentalist approaches. Narrative construction is proposed as the principle experience of self rather than a hypothesised 'real' self,

the self can be thought of as "distributed" among the individual and the whole group of others with whom he/she interacts. This distribution of identity is reflected in individuals' autobiographical narratives and practices (especially communication and information seeking and using' ((Lievrouw, 1996, p.223) citing (Gergen, 1982, 1991))

It is using this understanding of self that Lievrouw links research narratives with researcher identities. Research identities are fundamentally situational and 'contingent upon the changing fabric of a person's entire social experience and interactions with people'; narratives in effect constitute a researcher's dynamic identity (Lievrouw, 1996, p.223).

Lievrouw's (1996) articulation of research narratives and her propositional definition of its constitutive elements provide a scientific communication and practice based approach to conceptualising the importance of identity establishment and maintenance within the context of scientific work. Incorporation of understanding communicative practices in science has also been advocated by Gläser (2001b), who commensurate with Lievrouw's thesis, sees communication as having a specific role within science. Communication is presented as 'a strategic activity of actors' that is 'a means of coordinating these actions' and is 'a constituent part of [knowledge] production', having a role that significantly goes beyond exchange of information (Gläser, 2001b, p.194). Whitley (1969) also notes the particular importance of the communication system of science in the 'maintenance and growth of science'. It is 'the means by which the individual scientist relates to the social system' (p.219).

Gläser (2001a) builds on the conceptualisation of 'cognitive careers' (Mulkay, Gilbert, & Woolgar, 1975) to denote the role and participation of the scientist in the knowledge production of a specialty. This 'continuous participation' during a cognitive career produces an individual 'research trail' (Chubin & Connolly, 1982).

Throughout their 'cognitive careers' scientists accumulate knowledge by creating it and by memorizing knowledge, references to knowledge and references to knowledge retrieval procedures. This accumulation and creation process usually broadens scientists' knowledge base in the course of their career. Through this recombination and extension of knowledge, the cognitive careers provide the ground for scientific innovations. They allow new personal research plans to emerge that are based on new, individual recombinations of knowledge and permanently tested and adjusted in the scientist's research

work. These permanent processes of variation and selection are the main source of scientific innovations' (Gläser, 2001a, p.702).

Gläser (2001a) notes the lack of knowledge regarding the way aging influences this process of knowledge accumulation, recombination and the possible consequences for knowledge production (p.714).

Scholarly identities and other formulations of the mechanism of individual knowledge production, such as Gläser's conceptualisation of "cognitive careers", provide a foundational link between potentialities of citation identities and their relationship to the work practices of individuals. Lievrouw proposes that narratives may constitute the scientific work involved in identity development; if this is the case, she further argues that 'they can then be studied as indicators of how researchers negotiate, establish, and communicate their identities' (p.223). Selectivity, Ordering, Place, Ranking, Conflict, Closure and Interpretability are listed as being the narrative properties relevant to the research narrative.

3.3. SOCIOCOGNITIVE REPRESENTATION AND SCIENTIFIC WORK PRACTICES

Lievrouw introduces research narratives as mechanisms that are part of the communicative and representational scientific work practices of an individual researcher. Scholarly identities are constructed as dynamic outcomes of research narratives and involve both the communicating strategy of the researcher and its reception and relationship to knowledge production environments. This conceptualisation crucially incorporates the role of representation and a communicated 'identity' within the work practices of the individual researcher.

Representation as a scientific work practice that is a significant component of knowledge production has been studied extensively in sociology of science. Star (1995, p.98), citing the work of Latour and Woolgar (1979), and further exemplified in the work of Law and Williams (1982), writes 'in recent years much sociology of science has been busy documenting the gap between phenomena and representations'. The accounts within this programme of research are detailed analyses of the relationships between research process, scientific content and the situated feedback mechanisms for the communication of knowledge and its products. This section focuses on the work of Star (1983; 1995; 1989) and Courtial (2002) as a foregrounding to further discussion of the representational significance of Lievrouw's scholarly identities as a scientific work practice, as communicated through its presentational outcomes in knowledge products.

Following the work of Callon (1986) and Latour (1987; 1979), Star and Griesemer (1989) further articulate the processes underpinning the central role of representations in scientific work. Scientific work is fundamentally heterogeneous, and is 'conducted by extremely diverse groups of actors' (Star & Griesemer, 1989, p.387). However scientific work also 'requires cooperation – to create common understandings, to ensure reliability across domains and to gather information which retains its integrity across time, space and local contingencies' (p.387). The concurrent heterogeneity arising from 'divergent viewpoints' and 'the need for generalisable findings' creates a 'central tension' in science, which further requires processes of 'translation' between the diverse heterogeneous intersecting social worlds of participants in order to achieve coherence. Common representations become a central feature of scientific work practice, as

the creation of new scientific knowledge depends on communication as well as on creating new findings. But because these new objects and methods mean different things in different worlds, actors are faced with the task of reconciling these meanings if they wish to cooperate (Star & Griesemer, 1989, p.388)²⁴

One of the concepts Star and Griesemer (1989) introduce to explain how workers, in this case Museum workers, manage the tension produced by simultaneous 'diversity and cooperation' is 'boundary objects'²⁵. Boundary objects are scientific objects which 'both inhabit several intersecting social worlds ...and satisfy the informational requirements of each of them', they are 'plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites' (p.393). Boundary objects accommodate for different meanings in different social worlds but also have a common 'enough' structure to be recognisable and function as translations between social worlds. 'The creation and management of boundary objects is a key process in developing and maintaining coherence across intersecting social worlds' (p.393). The boundary objects presented in detail in Star and Griesemer's (1989) study of the Museum of Vertebrate Zoology include species and subspecies of birds, the terrain of the state of California and the habitats of collected animal species; however scientific publications and books are also listed as boundary objects in their definition (p.396). Creation of boundary objects is a specifically collective activity that emerges through the process of groups from different worlds working together, the intersectional nature of this work creates 'boundary

²⁴ This study deliberately stops short from the discussion of the epistemological concerns raised by Star in regard to the relationship between consensus and 'nature'. Meaning and consensus is specifically discussed in this work in terms of cooperative representation and its communicative significance in scientific work practices, with no position on further implied or explicit epistemological commitments.

²⁵ Van der Veer Martens (2001) further relates Star and Griesemer (Star & Griesemer, 1989) to Small's symbolic theory of citation [Section 2.8.1].

objects' which 'inhabit multiple worlds simultaneously' (p.408). By including scientific publications in this definition, they by association can be seen to incorporate these functional characteristics within scientific work practices.

Star's (1983) laboratory study of neuroscientists does specifically address the relationship between scientific work practices and publication outcomes. The premises with which Star begins this description of her study outline her approach to scientific publications as representations. Scientific work cannot be understood comprehensively through its products alone; however neither can it be understood without reference to these products. Scientific work 'involves joint effort over time, and thus is both interactive and processual', and 'meaning does not inhere in the nature of scientific work, but is continuously renegotiated by workers and consumers' (p.210).

Representation work is fundamentally embedded in scientific work practices and to some extent reflects the underlying research process; however scientific representations involve multiple simplifications that require more complete accounts of scientific work practices to consider both process and outcome. This does not mean that scientific products are not themselves worth studying. Rather, their representational and communicative role needs to be recognised as being a partial account that needs further analysis for it to be embedded within both the social and content aspects of the research process and scientific work practices more broadly.

Courtial (2002) underpins his approach to co-word analysis with a 'cultural' understanding of scientific articles as 'social representations in action'. This approach highlights both the content-dependant and social nature of scientific publications as representations of the social interactions of scientists. Scientific articles are regarded by Courtial as narratives where the 'researcher tells a story, puts actors on stage within a cultural script related to scientific culture' (p.222). As narratives, scientific articles 'take into account the implicit role of objects, the accessibility associations made by objects linked to ordinary scientific culture' (p.222). Through the presentation of narrative, scientific articles build links and thereby associations. These links 'form an image of the possible links between problems, the compilation of these links constituting what are generally known in the sociology of science as *problematic networks*' (p.222) [Courtial cites (Callon, Law, & Rip, 1986), author's italics]. Representations are translations from the research process and are dependent on the sub-culture and idiosyncrasies of the individual researcher; therefore representations are heterogeneous. They appear 'in a particular way, through analogies, paths which he introduces' that 'forge links between different problems' (p.223). Collective dynamics are highlighted 'by the systematic

analysis of interactions between researchers as detailed in their scientific publications' (p.223). A central feature of scientific representations is the associations between the author's problem and the problems of other researchers that are presented within them.²⁶

Star (1995; 1989) develops her analysis of the role of representations in situated knowledge production through the conceptualisation of 'knowledge ecologies'. Knowledge ecologies situate scientific work within layers of social organisations; they are inherently interactional across and within these layers, moving between individual knowledge producers, their immediate working environment, the organisational institutions they participate in, to the very broad social and political environment of funding policy and government. Star and Griesemer (1989) situate their scientists in institutional ecologies, the focus here is on organisational institutions; for example the Museum of Vertebrate Zoology which situates its knowledge workers within the institutional organisation of the Museum and then further into Berkeley University, as its academic institutional residence, and then further again into government and funding pressures brought about by interactions with macro level socio-political environments.

In this Museum (Star & Griesemer, 1989) study Joseph Grinnell's, the founding director of the Museum of Vertebrate Zoology, careful strategic alignment of organisational resources and interests, including field assistants, trappers, museum workers and the Museum's wealthy patroness is detailed. However Grinnell is clearly presented as further aligning all these other organisational resources and interests with the knowledge production interests and goal of contribution to the field of ecology. Quality and content of data collection throughout the whole organisational enterprise is guided by the strategic establishment and development of Grinnell's research programme, and the scientific problems that can be aligned with the organisational interests and resources of the Museum. This shift of emphasis from Grinnell as strategic organisational administrator and researcher (as it is presented in Star and Griesemer's account) to Grinnell as participant in a research collectivity, introduces consideration of Grinnell's placement of his research agenda within the problem areas

²⁶ Courtial uses co-word network analysis methods to empirically demonstrate dynamic associations in autism research. Co-word analyses have several limitations especially when words are derived from subject descriptor indexed fields, such as those used in this study using PSCHLIT. Citations on the other hand are author derived. Though word and citation data will provide different entry points into social representations [see Section 2.81], the conceptualisation that Courtial uses here is useful to consider in light of the specification of the role of representations in scientific work practices.

as established within his scientific community, and his membership within that community as guiding the content of the work he leads his organisational workers to do.

The importance of considering this shift of emphasis from Grinnell as organisational leader to research programme director and member of a scientific community is supported by the arguments of Gläser (2001b) and Lenoir (1997a), who establish content as being the key consideration driving social coordination in knowledge production. Lenoir argues (1997a) that for Latour and Woolgar (1979) arguments are ‘assembled by a stochastic process’ and while they are incorporated and accepted into scientific networks on the basis of ‘feedback’ from the broader scientific community, they are dependent on ‘chance’ (p.38). Latour and Woolgar’s account therefore does not provide for the institutional mechanisms that drive the content of science. In order to establish conditions for knowledge production that provide for the coordination of scientific content, scientific institutions that drive content and problem formulation across diverse research sites should be incorporated into accounts of representations in scientific work practices.

Gläser (2001a) locates the ‘formulation of research problems (i.e. the task), the selection of methods (i.e. the means) and the evaluation of the produced knowledge (i.e. the results)’ of individual researchers as within the ‘primary social loci’ of the specialty (p.702). Specialties and the participation of the researcher in them are said to define the nature of the scientific work of the researcher and its content. It is in the context of participation in specialties, either one or many, that ‘constitutes a scientist’s unique “cognitive career”’ (p.701). Specialties are defined by their common body of knowledge, which in turn through researchers’ continuous participation drive the content of the individual’s work; this is the means by which collective production is coordinated (Gläser, 2001b, pp.194-196).

Liewrouw’s (1996) conceptualisation of research narratives as communication mechanisms for establishing an individual researcher’s scholarly identity broadly addresses the individual’s placement within their scholarly communities. Her account specifies the social aspects of this communication but does not address the definitional role of content, and therefore the cognitive context within scholarly identity representational outcomes is addressed only in an auxiliary way. The following section outlines further conceptualisations of individual researchers and their representational contexts to attempt to address the considerations necessary to develop an account of

research identity and its representation that incorporates content and cognitive context.

3.4. KNOWLEDGE PRODUCTION INSTITUTIONS, INTELLECTUAL POSITIONING AND REPRESENTATIONAL SPACE

In partial response to the lack of support for the conditions of cognitive order found in Latour and Woolgar's (1979) account of fact construction, Whitley (1984) develops an extensive theory of the intellectual and social organisation of the sciences. His theory also addresses the issues raised in his critique of studies of the construction of knowledge that focus only on particular circumstances of knowledge production and do not contribute to developing 'comparative understanding of how different ways of conducting and validating research have become established and changed' (p.6). Whitley argues that 'comparative understanding is an essential part of any adequate sociology of knowledge which seeks to analyse how different knowledges are produced' (pp.5-6).

Science, for Whitley, is comprised of knowledge production units, which include 'university-based disciplines' (p.6). However, disciplines are not 'an essential feature of the modern science', and are historically variable as a basic unit of social organisation. Whitley introduces the 'broader and more general social unit of knowledge production and co-ordination' as the 'intellectual field'. Intellectual fields are defined as 'relatively well-bounded' and are,

distinct social organizations which control and direct the conduct of research on particular topics in different ways through the ability of their leaders to allocate rewards according to the merits of intellectual contributions (Whitley, 1984, p.7).

Though they are well-bounded, intellectual fields vary according the degree of their cohesion and autonomy from other social structures, 'but are the major social entities which co-ordinate and orient research across a wide variety of situations'; this orientation is around distinct subjects (p.7). Crucially for individual scientists, 'fields are the social contexts in which scientists develop distinctive competencies and research skills so that they make sense of their own actions in terms of these collective identities, goals and practices as mediated by leaders of employment organizations' (p.8).

Modern sciences are 'reputational systems' in Whitley's (1984) account. These reputational systems control and organise research through the 'institutional

commitment to novelty' (p.11). A scientist is rewarded most highly for scientific innovation that is further used to contribute to new innovations. Reputations provide the central mechanism for the organisation of scientific work; they are both reward and control for scientists. Colleagues need to be convinced of the novelty of the work, that the work has been produced using appropriate standard procedures and shared skills, and that the work is significant and relevant for the collective goals of the field. The production of new work is dually guided by the tension between novelty and tradition, or 'cooperation and competition' (p.13). Effective functioning of reputational systems of science require coupling with formal reporting systems or scientific communication systems, which 'enables task outcomes to be compared and coordinated' (p.19). Central to the coordination and control of the reputational system is the stratification of scientific work and the knowledge-producer of that work, according to the use by colleagues of the reported task outcomes of the work. Whitley's theoretical position produces an account for the systematic differentiation and characterisation of intellectual fields, according to the 'degree of mutual dependence between researchers' and the 'degree of task uncertainty in producing and evaluating knowledge claims' (p.85). Within Whitley's theory it can be seen that knowledge products and communication of task outcomes are crucially linked to the intellectual positioning and reputational success of the knowledge producer. Individuals are positioned within their intellectual field and are subject to their socially controlling influence through the acceptance and further use of their knowledge products by their colleagues. Individual researchers' placement, reputation and representation are defined in relation to their intellectual fields and they also contribute to defining the intellectual field's structure.

While Whitley (1984) is presenting a systematic and comprehensive account for the social control and content development of the sciences, that allows for comparative analysis of the differences in knowledge production conditions and their implications for accounting for the relationships between those difference conditions and the nature of the scientific work involved, there are features of his account that render it difficult to operationalise. For example, intellectual fields are defined broadly as knowledge units more general than disciplines; however at other times intellectual fields can be disciplines or even subdisciplinary knowledge production units. In order for systematic analysis to be undertaken according to Whitley's theoretical position, and typology of intellectual fields, these knowledge production units need to be bounded and therefore defined in an empirically meaningful way. Concepts such as mutual dependence and degree of task uncertainty need to be defined within intellectual field populations in

order for them to create comparative accounts of structure. The examples Whitley provides are primarily conceptual, and derived from historical understanding of the composition and features of the intellectual fields he highlights as exemplifying his typology, providing only limited solutions to the operationalisation problems raised.

Related to the definitional problem of intellectual fields is the identification of the locus of social control for reputational systems, and therefore an appropriate definition of what constitutes the subject and object of study of an intellectual field. As one among a multitude of alternative propositions available from the sociology of science Chubin (1976) delineates the teaching functions of ‘disciplines’ from the research functions of ‘specialties’. Chubin’s definition maintains the centrality of content, intellectual events and the relations they engender, as the defining feature of specialties from which social structures arise. Chubin also further develops the identification of specialties and their structures through ‘the nature of the communication relation used to link scientists’ (p.451). Gläser (2001b) extends Chubin’s discussion on the ability of scientometric techniques such as citation and co-citation analysis to contribute to definitions of specialty structures. Chubin is concerned with the difficulties in use of citation and co-citation measures for establishing specialty structures and identification, citing Mulkay’s “implicit theory of citing” and equating it with the assumption of “homogeneous” citing behaviour. Chubin however also advocates the use of these techniques for understanding specialty structures and conditions through their combination with intellectual histories of the specialties. Gläser (2001b) argues that communication links when established as representing use of a ‘common body of knowledge’, overcomes the need for reference to the homogeneity assumption. The definitional boundaries of specialties still remain; even with the introduction of membership as defined through participation, the heterogeneous and fractal nature of scientific structures still present difficulties for the delineation of knowledge production units’ populations.

The individual scientist forms an intersectional knowledge production unit across all other aggregated social and cognitive knowledge production units, and participation within these units by the individual is crucial for the production of new knowledge by the individual. The following sections introduce more detailed accounts of a selection of aspects of participation of individual within the conditions of their knowledge production units.

3.4.1. COMMUNICATION CHANNELS, PUBLICATION ACCEPTANCE AND INTELLECTUAL POSITIONING

Content and intellectual positioning happen within a field context that includes 'instruments of circulation, such as journals or publishing houses, which choose to publish articles and books in accordance with certain criteria and audiences in mind' (Lenoir, 1997b, p.11):

focus on scientific work entails examining the objective conditions that enter into the creation, circulation, and reproduction of the products of any given field. Within this perspective the author of a scientific text or theory is only the most visible node of a whole network of social relations, including authors of other scientific texts with whom he or she argues or from whom he or she draws support, publishers, instrument makers, lab assistants, university and state administrators...(p.10-11)

Whitley (1980) also comments on the complex role of communication channels (particularly journals); embedding them firmly in the recognition system of science, he is also explicitly relating them to the bounding activity of a field. Meaning and significance of research activity exists interrelated with recognition which locates 'particular research in particular intellectual contexts' (p.316).

Journals can be seen as means of institutionalizing intellectual commitments in that they reproduce in their own practices, and in their interrelations, the structure of norms and procedures which characterize particular areas of concern and work in the sciences (p.315)

thereby the 'selected outlet' will influence the 'particular topic' members within the field will choose to work on and the way they work These

intimate connections between journal publication and rewards and hence scientific 'careers' make the operation and organization of scientific journals key aspects of the organization of sciences in general, and of particular patterns of scientific investigation. Journals signify commitments to particular areas of concern and instantiate procedural norms. They thus both reflect the structure of scientific fields and reproduce them. (p.315)

Perspective audience has a crucial role within the knowledge production process through the acceptance and use of knowledge claims that are the 'currency' of individual researchers (Whitley, 1969).

Perceived individual identity and reputation have a role in the acceptance of individual scientists' work within broader institutional organisation (Zuckerman, 1978). In a study of peer-review and the refereeing process in science, Zuckerman and Merton (1971) found that 'unorthodox ideas' presented by 'established scientists' were more likely to be accepted than similarly 'unorthodox ideas' presented by 'young or rank-and-file'

scientists. Zuckerman (1978) extrapolates from this finding to argue that in practice theoretical ideas are responded to not necessarily according to either universalistic or particularistic criteria but that a 'special form of authority' [citing (Polanyi, 1958)] is being utilised 'which is based on scientists' previous records of contributions that produce differing degrees of confidence in their current work' (p.70).

3.4.2. RESEARCH MODELS AND THE CITING INDIVIDUAL RESEARCHER

Stets (2005) argues from the perspective of social identity theory that individuals act within the context of 'patterns of actions, interaction, and resource transfers' among all of the constitutive members of the structure of society. Individuals comprise, respond to and exist within these social structures. The self does emerge out of the mind; however concurrently the mind arises and develops out of social interaction, and patterned social interaction forms the basis of social structure (Stets & Burke, 2005, p.131). Negotiated meaning emerges from social interaction and identity. Behaviour is both guided by the relationship between situation and by internal self-meanings, allowing for both social structure and agency.

Gilbert (1976) locates negotiated meaning and identity in the functional attributes of 'research models' as they are embedded in his analysis of the processes of knowledge claim presentation and acceptance. It is 'models' that provide the underpinning of necessary relationships for Gilbert's positioning of the individual as central within the knowledge claim process. Models are defined as being 'an implicit metaphorical description of how some part of the world is thought to be arranged' (p.282); 'they can also be thought of as a system of interrelated concepts' so that concepts are interrelated within a model (p.283). 'Concepts are always obtained from well understood models, because otherwise they would have little meaning, but their meaning is modified as they become incorporated into the model' (p.283).

Models 'explain a very wide range of phenomena in terms of a single coherent system of concepts' (Gilbert, 1976, p.283)²⁷, but most crucially models belong to individual researchers. Models have functions similar to paradigms for the individual members of a field, but unlike Kuhn's paradigms, models can be shared though are not necessarily shared in entirety across fields or between individual members, so they are shared and

²⁷ This aspect of Gilbert's conceptualisation clearly requires further definition and empirical support; the central problem is that individuals are capable of holding multiple inconsistent concepts simultaneously. It can only be assumed that the research model conceptualisation produces the necessary reconciliation and coherence; however if this is indeed the assumption Gilbert is using, it is a very remote one.

differentiated by individuals within research traditions (p.301). An individual's research model has central functions throughout the research process, from establishing problems through to communicating the research knowledge claims through publications (p.284). The individual's research model is central to the way research is written, read and cited, as an individual's research model will be the reference point within or against which work is interpreted, integrated or rejected; the broad outlines of researchers' models are learnt in a tacit way 'as part of the process of becoming acquainted with work in their problem area' (p.294).

Researchers' models are the filter through which researchers relate to their problem areas and associated work practices (Gilbert, 1976). An individual's research model provides 'a initial formulation of the research problem, indicates the theories and techniques which may be applied appropriately to the problem, and eventually plays a crucial role in giving meaning to reports of the completed research' (p.302). Research models of others interact with the individual researcher's model through the assessment of the products of the individual's work at a communicated task outcome level; 'readers will extract from the reports those findings which suit their own models'. During write up citing authors

justify their findings by citing those conclusions which have gained a general acceptance within their field and which are also compatible with their own models. Certain findings will be repeatedly cited because they fit the majority of the models used by members of the network and can therefore become the basis on which research in the area relies for its justification (p.302).

The meaning of a knowledge claim is represented within publications by the contributing researcher/s, but those meanings will be then further interpreted by readers in terms of their own models. Individual researchers are able to be categorised 'into problem areas because the models they use bear a family resemblance to the models used by other working in the area' (Gilbert, 1976, p.297). "Family resemblance" is a deliberately loose definition as no two individual models within the problem area family will be identical; there are expected to be combinations of similarity and differentiation among individuals research models within research networks. Through this conceptualisation of individual research models, the individual's scientific work practices are directly associated with the researcher's role in knowledge production. This is further recognised as extending into the communication practices of the individual. The researcher's model relates them to their broader research network, and publications can be seen to be a crucial part of the work practices involved in generating and communicating knowledge claims. In this framework researchers as

individuals form a crucial component within understanding citation practice, this view incorporates individuals and their scientific work practices as they are embedded within research networks.

While Gilbert's individual research model framework is a promising theoretical foundation from which to examine the citing behaviour of individuals, his thesis has only been exposed to limited empirical testing. Gilbert (1976) includes only very broad empirical demonstration of the significance of the role of research models in the scientific work practices of individuals from his study on radar meteor researchers; this is only one group of researchers, and detailed articulation of the empirical implications are left relatively unresolved. Further empirical examination is warranted contingent upon appropriate operationalisation of potentially difficult concepts such as membership, the definition of research models as individual metaphors, and family resemblance.

Gilbert's (1976) account situates the individual in terms of their scientific work practices and places these practices within the individual's relationships to research traditions. However his account remains primarily theoretical with little further substantive demonstration through empirical evidence and specification of adequate operationalisation detail, which is required for a more generally applicable theoretical account to be developed.

3.4.3. RESEARCH TRAILS – THE ROLE OF PROBLEM CHOICE AND RESEARCH STRATEGIES IN REPRESENTATIONAL OUTCOMES²⁸

Chubin and Connolly (1982) conceptualise research trails as having a number of linking functions for the individual researcher as they work within their institutional knowledge production environments. Research trails represent 'a sequence of work by an individual or a small team of researchers' (p.295). This conceptualisation further leads to the definition of a specialty as 'the confluence of several research trails' (p.295). So the specialty becomes the aggregation site as composed by the 'choices of individual researchers' (p.294). The current composition of the specialty has implications for the choice of individual research trails and this choice 'is embedded in such issues as the distribution of research effort across problems in a specialty' (p.294).

²⁸ It is recognised that concepts represented here are consequential in establishing the importance of the role of the individual (and accompanying features such as careers) within the scientific enterprise. This requires longitudinal and qualitative methodologies to be addressed adequately, and cannot be addressed within the scope of this work.

A research trail thus directs attention to the coherence and development over time of a series or program of research projects undertaken by an individual or group (the latter often consisting of local colleagues), while the notion of a specialty directs attention to the coherence amongst several such trails at a particular point in time (Chubin & Connolly, 1982, p.295)

Establishment and maintenance of a research trail for individuals and groups requires their relationship to these institutional organisations to be recognised. This relationship is inherently reciprocal, and specialties are reconfigured in relation to the shifts in research trails and therefore individuals.

A slice through the bundle [specialty] at some point reveals some semblance of the membership and interrelationships between current members; following each wire from its source to its destination shows the linkages over time within one single trail (Chubin & Connolly, 1982, p.295)

An individual's intellectual positioning is determined by the constraints and order imposed from the organisation of specialties. Crucial career decisions for the researcher, such as problem choice and intellectual migration, are dependant on them. Conditions influencing research trails are both local to the individual or 'extra-local' and include legitimacy, funding, access to local resources and training capacity. Legitimacy and therefore the perception of the research trail is determined by 'definitions of acceptable novelty' (p.301) as developed through consensus at the specialty level (p.305). Specialties for Chubin and Connolly (1982) are defined by the aggregation of member research trails, in which the individual has a defining function; positioning of trails is socially negotiated and sensitive to local and extra-local pressures. However the definition of specialties is therefore inherently relative and dynamic, and the role of research trails in their sustenance is in aggregate and retrospective (p.304). This retrospective definition of specialty composition and the role of research trails within them lends itself to their examination from the perspective of the representational outcomes from these processes as operationalised through publication and citation use relationships.

Zuckerman and Cole (1994) report that prior research performance as assessed by peers 'differentiates among scientists and the ways they go about their work' (p.401). Zuckerman and Cole find that criteria of problem choice among 'eminent scientists' is differentiated from 'rank-and-file' scientists, through variation of significance of their work in terms of its potential reception by their field/s, extent of 'problem set' and choice of "strategic research materials" and "strategic research sites" or what to study. All these areas are related to the strategic placement of an individual and their work

within the broader context of their scientific environment.²⁹ 'The foci of attention in the sciences are of course no more than the aggregated problem choices of individuals working in those sciences' (Zuckerman, 1978, p.85). While Zuckerman here argues that problem area establishment is 'no more than' an aggregation of individuals' problems choices, Chubin and Connelly (1982) present the complex interrelationship between research trail development, problem choice and problem areas, that is more consistent with Zuckerman's previous statement that these choices co-exist with the communicative interrelations within those fields, such as the reactive nature of communication between theoretical and experimental contributions (Zuckerman, 1978, p.84).

Gieryn (1978) argues for examination of the issue of 'problem choice' from the orientation of scientific careers, as this provides a contextualisation of the problems chosen by individuals; 'a scientist rarely makes a career decision more consequential than the selection of a problem for research' (p.96). This argument is supported by a terminological typology of embeddedness of research choice, within problem areas, specialties and scientific disciplines, which necessarily relates the career significant (strategic) choices made by an individual to their broader scientific environment. An individual's problem set 'is defined as the set of problem areas in which an individual scientist does research at a designated time' (p.98). A number of problem formulations derive from the career orientation such as 'intellectual migration', 'inertial effects' of training, duration of research in a problem area, number of problem areas studied and kinds of changes in the sequence of problem areas.

While Gieryn (1978) is referring to processes of problem area choice he is actually defining these processes retrospectively by their operationalisation as 'publication of a scientific paper whose subject is within the substantive or technical scope of a problem area' (p.98). This suggests that artefacts such as publications reflect outcomes of problem choice decisions. Social representations through formal publications cannot directly access the decision making and motivational processes of the individual researcher; however they can indicate outcomes in retrospect.

Scholarly identities and their constitutive research narratives are in part contingent on the processes of research trail development and problem area choice. Lievrouw (1996, p.221) argues that researchers define the terms, boundaries and specificity of their

²⁹ Note that the scientists selected for this study were predominantly from the physical and biological sciences and mathematics, and relatively few from psychology and economics.

research through the ‘extraordinary’ selectivity demonstrated in their choice of problems. ‘Narrative construction/identity formation influences, even drives, the types of problems a scholar chooses to explore and the answers she[/he] is likely to formulate for those problems’ (p.229). Viewing artefacts as one of the outcomes of social representations of knowledge construction suggests that researcher identity, as examined through citation identities (as outcomes of broader identity formation), will be represented through source use by the individual researcher, where shifts and problem attention can be examined retrospectively.

3.4.4. MEMBERSHIP, MARGINALITY AND MIGRATION

The sociocognitive positioning of an individual, their scientific work and the representational outcomes of this work is relational to the knowledge production communities to which individual researchers belong. Membership within these communities is crucial in establishing the placement of the representational outcomes of individuals.

Gieryn and Hirsh (1983) examine the hypothesis that ‘innovations are produced more often by scientists at the margin of a field than by those closer to the centre’ (p.87). They find that marginality seems to be less important than ‘the juxtaposition of old and new ideas or techniques introduced by the migrating scientist’ for the occurrence of innovation (p.100). Marginality and ‘intellectual migration’ are features of individual human actors as they are placed within and related to their broader fields.

Kuhn (1963) identifies the significant role of balance between tradition and innovation in knowledge production, he calls this the ‘essential tension’. For Kuhn, ‘only investigations firmly rooted in the contemporary scientific tradition are likely to break that tradition and give rise to a new one’ (p.343). Scientific creativity is predicated on this balance at a group and individual level. To be successful, scientists ‘must simultaneously display the characteristics of the traditionalist and of the iconoclast’ (p.343). The success of this venture is determined by the group, and group membership consists of varying degrees of balance. In order to empirically examine this relationship between tradition and innovation and therefore the mechanisms of knowledge generation, as hypothesised by Kuhn, it is necessary to establish the ‘tradition’ that is being referred to, so that breaks from that tradition can be identified. This is not a straightforward task; however as Chubin and Connolly’s (1982) conceptualisation of specialties and their relationship to research trails supports, group membership provides a basis for such definitions and that individuals, through the interaction,

selection and membership choices they make are the defining features of research 'tradition' development.

Star and Griesemer (1989) draw an analogy between 'immutable mobiles', 'objects which can be transported over a long distance and convey unchanging information', and 'marginal people'. Marginal people refers to 'a person who has membership in more than one social world' (p.411). Marginality is seen by Star and Griesemer as a 'critical concept for understanding the ways in which the boundaries of social worlds are constructed, and the kinds of navigation and articulation performed by those with multiple memberships' (p.411). The strategies that marginal people employ to manage their identities is seen as a 'provocative source of metaphors for understanding objects with multiple memberships' (p.411). While strategies themselves are not articulated directly within the representations created by marginal people, examination of the outcomes of those strategies via the publication and referencing patterns of individual researchers, such as those produced in citation identities and intercitation, may inform studies of the individual's relationship to their sociocognitive networks and the outcomes of their placement.

In his work on scientific cultures Becher (1989) refers to individual academics as the 'elementary particles of the intellectual world' (p.105); these elementary particles exist and function within a complex array of larger social and intellectual communities. Academic 'territories' are comprised of overlapping and interrelated layers of community. Becher emphasises the diverse range of forms which broad knowledge fields, disciplines and specialisms take and the difficulties in identifying boundaries of these socio-cognitive units. He conceptualises specialisms as forming into the larger groupings of disciplines but the relationships amongst the varying levels of grouping is difficult to disentangle, primarily due to the inter- and intra-level convergence and divergence of memberships and practice. All levels of 'territorial' groupings demonstrate this specialisation and unification activity in an infinite variety of ways.

Individual academics as the elementary particles within this activity themselves need to negotiate this complex socio-cognitive environment via yet another sub-level of interrelationships that immediately have implications for their scientific work practices. Academic careers and intellectual identities (Becher, 1989, p.111) are created and expressed through nesting and overlapping specialisation of memberships within and across the broader groupings of science, which provide 'frames of reference' (p.49) for many aspects of the individuals work practice. Becher considers the definition of

individuals in relation to their specialism must be achieved through establishing the 'labels' scientists accord themselves (p.44).

Beyond specialisms it can be seen that multiple levels of participation are in evidence during the process of producing knowledge. The examination of researcher identity as reflected in the formal communication artefacts of this process holds potential for contributing to understanding an individual's relationships to their various communities through the prominent role of artefacts in knowledge production and communication processes.

Citations and publications are outcomes produced during knowledge production processes. Though it becomes apparent through examining the literature on citation that there are complex relationships to be considered within citation practices and the formal outcomes of scientific communication, a form of self-labelling is manifest albeit not in a straightforward clear statement of participation. Whitley (1984) recognises referencing as a 'major manifestation' of social control in science and so in a restricted sense references are a demonstration of 'group membership and identity'. Connection of new work to currently existing evidence ensures 'work is not too far removed from the aims and procedures of the dominant group' and 'the degree of innovation is thus diminished and constrained by the necessity of showing how new contributions fit in with, and are relevant to, existing knowledge' (p.28).

Membership and group identity remains a central problem in citation analysis and its extension into analysis of the placement of the individual within larger knowledge production units. Use of individual sources and knowledge claims is not delineated across or within specialty or disciplinary groupings, as found by Leydesdorff and Amsterdamska (1990) in a questionnaire on individual researchers' specialty and disciplinary identity referring to the use of four individual publications. The question of specialty membership was addressed by Laudel (2003) in her examination of the spatial migration of specialty elite. This study addressed multiple concerns that are beyond the scope of this work; however spatial migration and its establishment has a number of similar concerns as intellectual migration, and by extension intellectual membership and placement. In order to establish intellectual migration, conceptual change must be identifiable which further involves movements in the intellectual placement of an individual and their work in relation to a reference 'group'. Again referring to Chubin and Connelly's research trails, individuals are located within specialties as primary organisational intellectual and social units and these units are

defined by participation through research trails. Laudel's (2003) operationalisation of specialties through a methodology that combines aggregated co-citation, citation counts, biographical data and participation in significant specialty events allows for a reference group of individual members to be established, which goes some way to developing a context for individuals' sociocognitive placement in relation to other individual members of an intellectually and socially significant reference group.³⁰

3.5. SCOPE OF THESIS

The above review has encompassed issues of knowledge production units and field placement, to give a broad representation of studies on identity and the communicative work practices of individuals in knowledge production. The establishment of a researcher's identity is shown to be a multi-faceted problem. The establishment of scholarly identities provides, the ability to characterise and differentiate researchers. The process of communication is central to this establishment of scholarly identity. While focused on narrative, scholarly communication draws highly on the use of citations to establish intellectual position, therefore it is feasible to consider the analysis of citations to inform or embellish the scholarly identity.

Preliminary establishment of the notion of citation identity has been undertaken by White. The establishment of a methodology, and the understanding of the broad applicability of the citation identity, are in preliminary stages. However, the prospect of determining characteristics of characteristics of a researcher through interpretation of citation patterns is interesting for providing a separate perspective on identity to other sociological and citation analysis methodologies currently in use. Furthermore, the sociological processes require intense personal interaction, whereas the use of citation offers the possibility of a complimentary systematic examination, leading to automated examination of field through identifiable citation characteristics.

In understanding identity, it is clear that citation analysis will never provide the rich picture of other forms of analysis; therefore research in this area should be based on determining where citation analysis can provide complementary information, as well as confirmation, to the understanding of identity. Additionally, any attempt to address identity through citation analysis will have to successfully place the interpretation of identity within these broader frameworks: membership of group, field and institution;

³⁰ Laudel's (2003) methodology is further described in relation to the identification of individual researchers and their placement in the field of Consciousness in Chapter 4.

problem choice and career direction; placement within communication systems, and the development of reputation.

The scope of thesis is exploration of use of the citation pattern of the individual, in an effort to develop the means of addressing scholarly identity as a representational scientific work practice through the use of citation analysis. The research will be based on an expanded study of authors from diverse backgrounds, to provide an understanding of the ultimate applicability and likely success of the analysis to informing scholarly identity. A range of citation analysis methods will be used to compare the use of citation identities with broader analysis of field, to address questions of the future determination of scholarly identity through citation analysis.

4. CITEE-CITATION CONCENTRATION AND CITING IDENTITY TYPES

4.1. INTRODUCTION

A significant challenge identified in the review of literature on representation of the individual (Chapters 3) is the adequate representation of the individual through citation analysis. The challenge of characterising researchers is, in theory, made easier by the differentiation of authors according to a citing identity, where that citing identity can be adequately related to a set of identifying features. This concept has had limited development to date, and has been applied only in small fields. A limitation of definition of citing identity is the ability to establish a consistent platform for comparison. This becomes more acceptable with verification of methodology across diverse fields, where a suitably common baseline can be established.

The current chapter will explore the use of a new methodology for the generation of citing identity types, based on the identification of consistencies between authors' citing practices. Researchers working on the problem of consciousness are selected as the subject of study, to provide a broad basis for establishing this new methodology.

4.2. THE SCIENTIFIC STUDY OF CONSCIOUSNESS

4.2.1. INTRODUCTION TO THE SCIENTIFIC STUDY OF CONSCIOUSNESS

The field of consciousness is currently in its infancy. The conference *Toward a Science of Consciousness* held in 1994 is the first instance where a diverse range of individual researchers were brought together in a recognisable formal group in order to define the 'problem of consciousness' (Hameroff, 1994). This initial conference and its subsequent series of conferences were established with the specific aim of pursuing the 'scientific study of consciousness' (*Toward a Science of Consciousness*, 2000; *Toward a Science of Consciousness: The First Tucson Discussions and Debates*, 1994; *Toward a Science of Consciousness: The Second Tucson Discussions and Debates*, 1996; *Toward a Science of Consciousness: The Third Tucson Discussions and Debates*, 1998). Previous to this occasion the scientific study of consciousness only existed as it was distributed under the auspices of a number of divergent disciplines, including Philosophy of Mind, Cognitive Science, Psychology, Medicine and Artificial Intelligence. The working party for the first conference and the first editorial meeting for the *Journal of Consciousness Studies* took place in 1993 and consisted of similar members (Editorial: the future of

consciousness studies, 1997). It is also significant to note here that the Editors for the *Journal of Consciousness Studies* are approximately the same group of researchers as those maintaining the conference series (ASSC: Association for the Scientific Studies of Consciousness, No date; *Journal of Consciousness Studies*, no date).

There is continuing debate in the field as to the relevance of contributing disciplines (Petranker, 2001). The Editors of the *Journal of Consciousness Studies* (Editorial: the future of consciousness studies, 1997) provide the following list of the ‘kosher’ and ‘taboo’ disciplines:

<i>Kosher</i>	<i>Taboo</i>
Philosophy	Botany
Neurobiology	Hermeneutics
Cognitive science	Healing
Physics	Literature
Phenomenology	Folk psychology
Sociology	Aesthetics
Anthropology	Anomalies
Feminism	Religion
Ecology	Psychotherapy
	Ethics
	Transpersonal psychology

This categorised list is argued to be representative of opinion in the field; however the editors then proceed to argue that they do not consider this categorisation in their journal selection decisions, as the distinction is ‘wrong in principle’ for the pursuit of consciousness as a scientific study for the following reasons:

1. No-one has as yet come up with any evidence for a theory of consciousness that will satisfy the demands of the various sceptics, so the decision to focus the investigation at, say, the level of the neuronal network has to be for pre-theoretical reasons.
2. We only know consciousness through our own experience, so arguments against including a first-person phenomenological approach are contradictions in terms.
3. The only form of consciousness that we know directly is human, and this is characteristically shaped by social, cultural and environmental factors.

The editors stated policy for inclusionary selection decisions for the ‘fields’ journal and the conference series is argued to be not effectively applied in practice (Sutherland, 1998; Whitehead, 2004). Sutherland and Whitehead both see a growing trend toward

over-representation of neuroscientists and cognitive scientists to the exclusion of anthropologists and generally ‘alternative approaches’. Decisions as to fields of study to include are accompanied by continuing negotiation of what Consciousness is as an object of study (de Quincey, 2006).

To the knowledge of the author, the field of consciousness has not been studied previously using bibliometric analyses. Studies have been completed in a number of contributing fields, including Neuroscience (Cromby, 2004; Schwechheimer & Winterhager, 2001), Neural Networks (McCain, 1998), Artificial Intelligence (van den Besselaar & Leydesdorff, 1996) and two sub-areas of Philosophy, philosophy of science and epistemology (Kreuzman, 2001).

4.2.2. THE SCIENTIFIC STUDY OF CONSCIOUSNESS AS BOUNDARY SPECIALTY

As can be seen from the editorials and field commentaries outlined above, the scientific study of consciousness is a problem area that has had a very particular beginning. Establishment of the field has been orchestrated through the bringing together of researchers from deliberately diverse backgrounds with broad based concerns in the general area of the ‘problem of consciousness’. The field is relatively young in that it has only formally received attention as a field from 1993, when researchers began attempts to establish it, through inception of the *Journal of Consciousness Studies*, the *Association for the Scientific Studies of Consciousness (ASSC)*, and the *Toward a Science of Consciousness* conference series.

Further work is required to define the nature of the field of consciousness in terms of its formal definition as a specialty (Chubin, 1976), problem area (Gilbert, 1976; Mulkay, Gilbert, & Woolgar, 1975), problematic network (Courtial, 2002) or intellectual field (Whitley, 1984) from the definitions provided by the literature. This work is beyond the scope of this thesis, and is considered for the purposes of the work presented, though desirable, unnecessary. The object of study for this thesis is the individual knowledge-producer and their representation through formal publications. However, as argued in the previous chapter, this requires recognition of their sociocognitive placement within their broader knowledge production units. In order to establish the analysis of the individual knowledge-producer within their representational space, the definition of the scientific study of consciousness as a boundary specialty is introduced. It is not claimed that the field of consciousness is a specialty by formal definition; however using a weak interpretation of Gläser’s (2001) definition of specialty as defined by a ‘common body of knowledge’ and members as defined as ‘continuous participants’, the

identification of consciousness used in this study complies. Boundary is used as derived from Star's definition of 'boundary objects'. As outlined in Section 3.3, Star and Griesemer (1989, p.393) describe boundary objects as being 'plastic enough' to accommodate for different meanings in different social worlds as well as being 'robust enough to maintain identity across sites'. In the case of the scientific study of consciousness the problem of consciousness itself can be seen to be an instance of such a boundary object for a diverse range of researchers.

4.2.3. ESTABLISHING RESEARCHER PARTICIPATION – DEFINING THE DATA SET

Contributing authors were identified through a combination of participation lists developed from the *Toward a Science of Consciousness* conference series (*Toward a Science of Consciousness*, 2000; *Toward a Science of Consciousness: The First Tucson Discussions and Debates*, 1994; *Toward a Science of Consciousness: The Second Tucson Discussions and Debates*, 1996; *Toward a Science of Consciousness: The Third Tucson Discussions and Debates*, 1998) and subject searches in the Institute for Scientific Information (ISI) citation indexes available through DIALOG.³¹ The conference series is significant as it was initiated and developed in order to establish the problem of consciousness scientifically; it deliberately brings together researchers that in some way contribute to work on the problem of consciousness and are selected from a deliberately formed base of researchers with diverse disciplinary affiliations.

The following lists were developed to identify Consciousness boundary specialty contributors (all lists were standardised for name variants):

- Cited author (CA) list – ISI citation indexes

Search of the word Consciousness in the Identifier (ID) field combined with a combination of the Descriptor (DE) and Title (TI) field. In the ISI citation indexes (Dialog files 7, 34, 434, 493). Duplicates were then removed (using the remove duplicate command (RD)) followed by list developed using Dialog's RANK command.

- Author (AU) list – ISI citation indexes

Search of the word Consciousness in the Identifier (ID) field combined with a combination of the Descriptor (DE) and Title (TI) field. In the ISI citation

³¹ File 7: Social SciSearch(R) 1972-2001/Mar W2 – Social Sciences Citation Index
File 34: SciSearch(R) Cited Ref Sci 1990-2001/Mar W2 – Science Citation Index
File 434: SciSearch(R) Cited Ref Sci 1974-1989/Dec – Science Citation Index
File 439: Arts&Humanities Search(R) 1980-2001/Mar W2 – Arts and Humanities Citation Index

indexes (Dialog files 7, 34, 434, 493). Duplicates were then removed (using the remove duplicate command (RD)) followed by list developed using Dialog's RANK command.

- Conference series cited author list

List of cited authors extracted from conference proceedings of Science of Consciousness studies 1994 and 1996, and ranked according to number of times author cited. This list is not restricted to first cited author only.

- Conference series author list

List of authors extracted from conference proceedings of Science of Consciousness studies 1994, 1996 and 1998, and ranked according to number of times author cited.

These lists were then compared and authors that appeared across the conference author list and the conference cited author list were included. Then authors from the ISI cited author list and conference cited author list were included. Authors that were ranked eight or above in the cited authors counts from the ISI cited author lists were also included. The resulting list of conference and boundary specialty participants comprised 128 individuals.

An author co-citation matrix was then developed (using techniques derived from (McCain, 1990)). The years included for the matrix were 1990 to 1999, the search was performed in 2001, and from the conference proceedings was intended to capture the first ten years of the boundary specialty as it was represented in the conference organisers' opinion. This search strategy limits the time of citing publications, not the time of the cited publications. Searches were again performed in the ISI citation indexes, and so have the limitation of only retrieving the first cited author for any one cited reference. Variations identified in the initial development of participant and author lists were used to retrieve maximal comprehensivity.

Sample search for co-citation matrix data collection:

```
S CA=(AHERN GL OR AHERN G)/1990:1999
S CA=(ALKIRE MT OR ALKIRE M)/1990:1999
S1 AND S2 NOT UD=2001?;RD
```

Co-citation counts for each author from the matrix were then ranked; the three authors that were not co-cited with any of the authors from the matrix were removed from the set. From this list of 125 authors, 25 authors were removed. The point of cut-off was pragmatic in that SPSS only allowed for 100 variables, this number of authors

coincided with a 'step' from authors being co-cited more than 100 times to those being cited less than this.^{32,33}

4.2.3.1. GENERAL LIMITATIONS WHEN USING ISI CITATION INDEX DATA

The methodological reliance on the ISI citation indexes for co-citation analyses means that the resulting comprehensiveness of the analysis is limited by the scope and composition of the databases. Notable problems include selection criteria for inclusion of journals indexed and nationality biases inherent in the USA production and therefore English language bias of database coverage. Finally, with journals being the primary source material of the ISI databases, biases are inherent through the publication patterns of individual boundary specialties, for example, the well documented preference toward communication through monograph literature of the social sciences and particularly the arts and humanities.

4.2.4. COLLECTION OF BIOGRAPHICAL AUTHOR INFORMATION

Biographical information for participating consciousness authors was obtained through iterative searching of the internet, through a combination of author name searches, searches of university sites (unreliably available from the Corporate Source (CS) field in the citation indexes) and library catalogue searches for PhD titles (again where other information had been obtained that allowed this). Where possible, Curriculum Vitae (CV) data was obtained. However, as noted by Laudel (2003), curriculum vitae information obtained in this way is dependant on the behaviour of the individual researcher and is therefore not consistently available. In total only 15 full CVs from the 100 author set were obtained. However other verified and detailed biographical was available, for example through Nobel Laureate biographies and some University department or research organisation websites.³⁴ Unlike Laudel's study, biographical information was used only for background purposes and comprehensivity, though desirable, was not so crucial to the analysis. Finally in cases where basic biographical information such as date of birth or PhD qualification was not available, scientific biographies were searched online³⁵ and in hardcopy. Overall, key biographical dates

³² It is recognised that this is an essentially arbitrary threshold for the identification of a specialty or intellectual field population. However the following analyses using aggregations of participants in the field seek to describe the relationships between participants within their representational space, rather than claim to be structural analyses of the field as a whole.

³³ See Appendix A for a list of authors included in the author co-citation analysis

³⁴ Translation of Giuseppe Moruzzi's biographical information from an Italian University was supplied by Guido Orsatti

³⁵ Marquis Who's Who – Dialog file 234

Bowker Biographical Directory – Dialog file 236

such as date of birth and year of highest degree completion were obtained for all the authors, though many other career details are incomplete.

4.2.5. COLLECTING CITING IDENTITIES

The 100 author list resulting from the co-citation counts was further reduced by three authors. The excluded authors WILDER PENFIELD (1891-1976), SIGMUND FREUD (1856-1939) and WILLIAM JAMES (1842-1910), though able to be adequately captured using co-citation searches for the purpose of author co-citation analysis, were not adequately represented by the ISI citation indexes for citing identity development. As can be seen two of the authors, SIGMUND FREUD and WILLIAM JAMES, died sometime before the earliest date for the beginning of the ISI citation indexes (1972). The third author, WILDER PENFIELD, died only 4 years after coverage begins. The contributing author list for citing identities was stabilised at 97 authors.³⁶

The ISI citation indexes were then searched again using the searches developed for the co-citation matrix for these 97 authors, and duplicates were removed using the RD command.³⁷ The searches were for all available years prior to and including 2002. The searches were not further limited in order to capture potential author publication records as comprehensively as possible. Publications that listed the author in any position in the by-line were included. Retrieved records from these searches were manually verified using biographical data (section 4.2.4) and publicly available author produced publication lists where possible. In cases of authorship doubt, full-text articles were obtained and author summary details cross-checked with biographical sources. In 47 instances records that had been verified as belonging to a contributing author had “no keyed refs”, meaning that cited reference data had not been included in the record. For these records full-text of the article was obtained and the references scanned and included following the conventions used by records in the citation indexes.

Once records were verified as being the contribution of the target author, the cited reference (CR) field was extracted using a purpose written program [see Appendix

Dissertation Abstracts Online – Dialog file 34

³⁶ For the complete list of authors developed for citing identity analysis and author co-citation matrix see Appendix A.

³⁷ The collection of citing identities used in this study differs from White (2001) in two ways. Searches are not limited to first or sole contributing author, as it has been argued that name ordering on publications is heterogeneous with regard to authoring role (Harsanyi, 1993; Vinkler, 1993; Zuckerman, 1968); though Laudel (2001) demonstrates that first author position indicates that the named scientists did ‘substantive creative work’, the amount of it cannot be assumed from the author list alone. Furthermore the search functionality of DIALOG is not relied on so that publication and author tracking can be more directly monitored from the verified record set.

B.1].³⁸ A unique identifier was assigned to each record to allow for verification and tracking of cited references according to the publication and its author. Once extracted, cited references were further collated and analysed using Excel. Further verification and data standardisation was needed to, as much as possible, ensure consistency and correct homonym concatenation, for cited references. From these ‘raw’ individual author cited reference lists, cited author lists and counts were ranked and Gini and citee-citation measures calculated as presented in the following sections of this chapter.

4.3. GINI AS A MEASURE OF CITEE-CITATION CONCENTRATION³⁹

In order to examine the citee usage of individual authors more descriptively according to the diversity of distribution within an author’s citing behaviour (or White’s “citation identity”) the Gini measure of concentration or diversity (or inequality) was introduced.

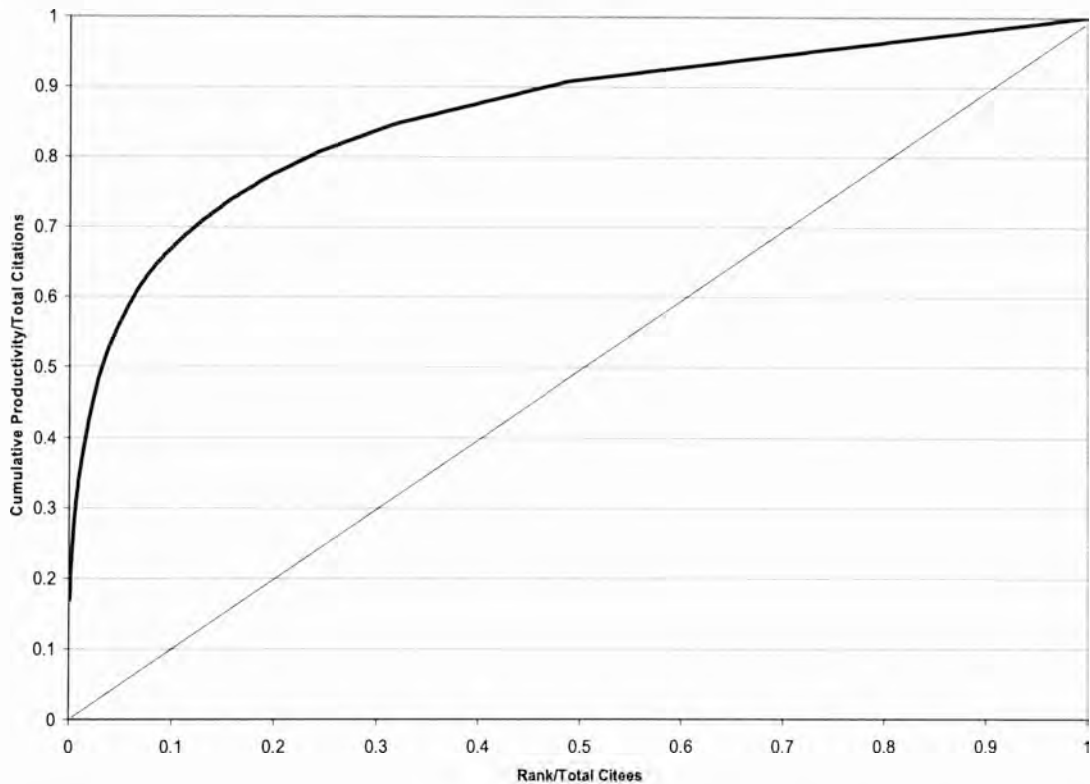


Figure 4.1 Example of an individual author’s normalised cumulative citation and citee distribution

Figure 4.1 contains a plot of normalised cumulative citations and citees, produced by an individual author’s entire publication career (in this case as represented in the ISI citation indexes). This distribution can be characterised in different ways. It can be

³⁸ This program was co-developed with Andrew Beehag
³⁹ In this chapter the convention of referring to the citee-citation ratio established by White is maintained. Chapter 5 introduces the terminological shift required to differentiate citations from references (referred to as cites).

seen that the curve begins some way off the y-axis; this point is the most highly cited citee, usually consisting of self-citations. The slope of the line from this initial point characterises the more highly cited authors. The right hand region of the graph is a shallow sloped straight line indicating unicitations. The elements indicated here are patterns observed across all authors' citation oeuvres, which have these characteristics in some form. The diagonal in Figure 4.1 represents where the distribution would lie if all citees received the same number of citations.

The Gini coefficient of inequality was used in an attempt to characterise the different aspects of the individual authors' citee-citation curve. If all citees received the same number of citations, as indicated by the diagonal line, then the Gini would be equal to zero. In the case where the Gini approaches one, the author will be characterised by the first citee having the vast majority of total citations with other citees making up the remaining minority. For example, an author with a high proportion of unicitations will have a curve that is close to the diagonal line therefore the Gini value is low. Conversely, an author with a core of highly cited authors and fewer unicitations will have a curve that bulges away from the diagonal, and therefore the Gini value is high.

The use of Gini as a measure of citation inequality and concentration has some precedence in the informetric literature, for example Stegmann and Grohmann (2001) use the Gini measure to determine concentration of citation to journals as part of a suite of measures aimed at capturing journal visibility. Burrell (1991) also advocates using the empirical Gini as a comparative numerical measure of the degree of concentration generally in the field of bibliometric studies. His investigation concerns bibliographies of fields of study, which to some extent is analogous to the study of citation from individual authors or White's (2001) "citation identities", in that both are examining references from, as opposed to citations to, a field or individual.

The range of Gini values found within the 97 consciousness researchers studied is IAN MARSHALL at a Gini value of 0.15, the lowest, to THOMAS NATSOULAS with a Gini value of 0.77, the highest within the set.

4.4. THE GINI VERSUS CITEE-CITATION RELATION

White (2001) introduced the citation-citee ratio, as an application of the type-token ratio (Herdan, 1960), to look at citation styles of individual authors. In this study the use of this measure is further extended by applying the Gini measure of inequality, to further capture characteristics within an individual's citation patterns. The comparison

of the two measures of citation, applied to citees and their relationship to citations within an individual's citation oeuvre, is shown in Figure 4.2.

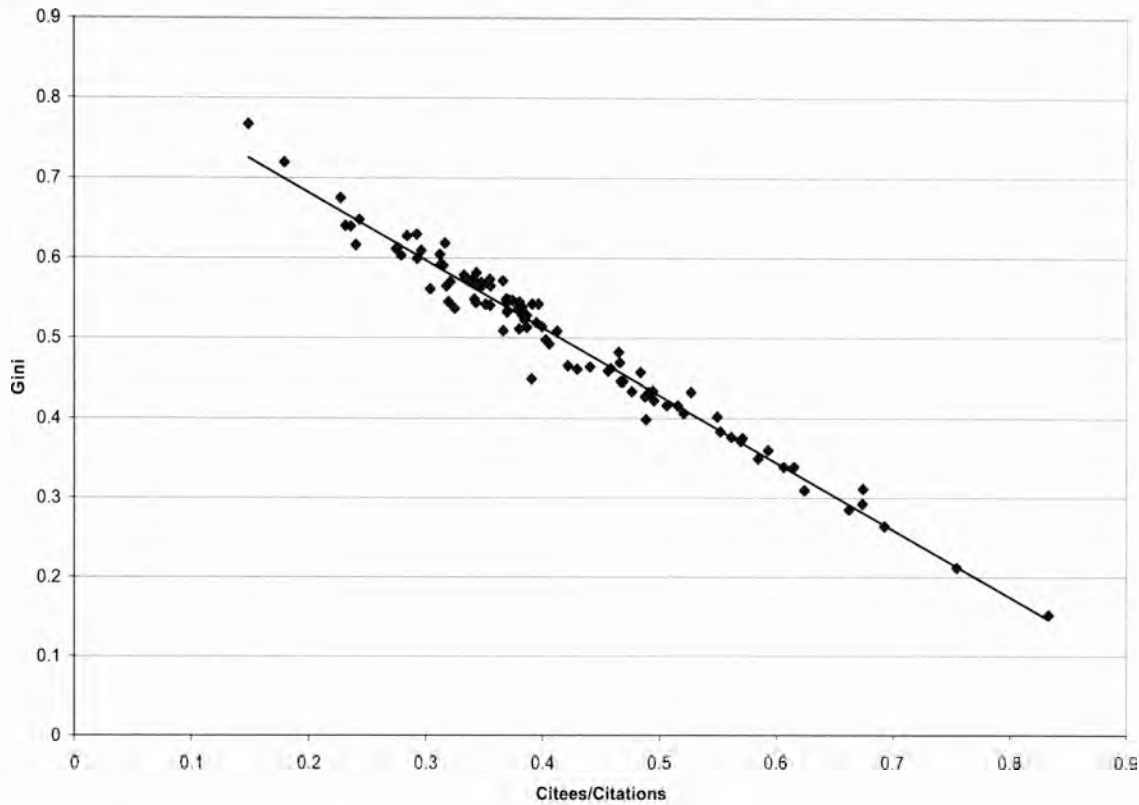


Figure 4.2 Gini versus Citee-citation ratio

When the Gini and the citee-citation [type-token] relation is calculated for the 97 consciousness researchers in the sample and regressed against one another, a correlation of $R^2 = 0.972$ is found. This high correlation is in keeping with other findings using type-token ratio measures, covering a variety of phenomena in a variety of formulations. White (2001, p.94) finds a R^2 of 0.94 within his analysis of 8 information scientists. The present study of 97 consciousness researchers uses Herdan's (1960, p.30) original formulation of the Type-token ratio, the citee-citation ratio. A calculation of correlation using White's formulation of the citation-citee ratio applied to this data set results in an R^2 value of 0.90. Losee (1996, p.753) uses the type-token ratio to examine use of terminology in four different databases; data reconstructed from his averages of type and tokens from the full text, result in an R^2 value of 0.95, though it is noted that this is a very limited sample size as only averages are given. Herdan finds a similarly high correlation in his analysis of vocabulary occurrence in two sections of the Bible using the bilogarithmic type-token ratio, The Pauline Epistles, and a number of sections of the New Testament. The logarithmic power law correlations are 0.95 and 0.80, respectively (Herdan, 1960, pp.30-31).

Some correlation was expected between the citee-citation ratio and the Gini; as an example, when the citee-citation ratio approaches one, there is nearly an even distribution of citations for each citee, which is the same condition for defining a Gini equal to zero. However for any further correlation, below citee-citation ratio values close to one, it would be expected that individual authors would need to cite across a number of characteristics in a common way. For example, their proportion of first cited author to uncitation authors would be consistent for an equivalent citee-citation ratio. But as can be seen with the citation identities of authors ALAN COWEY and JOHN HOBSON, both at an equivalent 0.28 and 0.31 citee-citation ratio, have nearly equivalent total citations, 6159 and 6510, and nearly equivalent amounts of recitations, 5224 and 5284, but they also have substantially different highest cited author counts, 216 and 407 (in both cases self-citations), as well as somewhat different uncitation counts, 945 and 1262. Given these differences it would not be expected that these two authors would have similar Gini values; however their Gini values are very near equivalent at, 0.60 each.

The Gini is an accepted measure of concentration within a distribution, whereas the cite-citation ratio is by comparison a relatively simple measure. Despite this, the results of comparison of the two measures in this analysis indicate, with citee-citation data, that the two measures can be equated.

For purposes of comparison, a plot of the Gini versus log citee over log citation was generated. The bilogarithmic type-token ratio was the original formulation by Herdan (1960, p.28), where it is used to develop individual author stylistics, comparable within and across texts. Herdan's (1960) argument is that the type-token ratio in non-logarithmic form is unable to compare author stylistics across texts as it does not account for variation in text length; he converted it into bilogarithmic form to overcome this limitation. When the Gini is compared with the bilogarithmic form of the citee-citation ratio a far more approximate relation is found, as can be seen in Figure 4.3, a linear fit produces an R^2 value of 0.67 whereas a second order polynomial fit increases this slightly resulting in a R^2 value of 0.71.

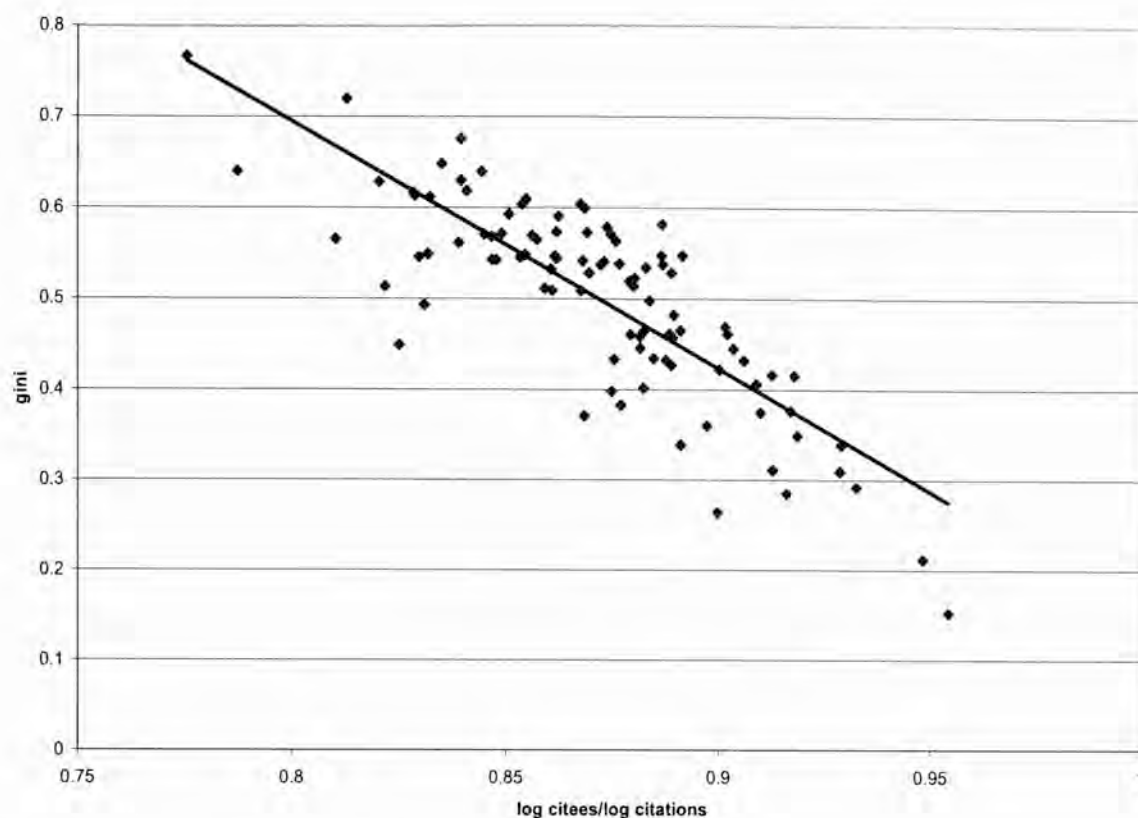


Figure 4.3 Gini versus log citee-log citation relation

It is expected that there would be a difference when comparing the log citee-log citation form to the Gini as the log citee-log citation is different to citee-citation; however the R^2 is significantly lower in the bilogarithmic analysis. For instance the comparison of Gini and the log citee-log-citation ratio did not result in a logarithmic fit. This has implications for the establishment of citation behaviour types. Noting that Herdan (1960) advocates the use of the bilogarithmic form of the type-token ratio (in this case the citee-citation ratio), each of the variations of citation behaviour measures can be discussed.

If it is accepted that the bilogarithmic citee-citation form is the best indicator of citation behaviour type, then it would appear from the above result that the Gini is not a good measure. At a Gini value of 0.58 the value of the bilogarithmic citee-citation ratio varies from approximately 0.81 to 0.885, or 40% of the spread of the entire data set. However, the converse consideration is that the Gini is a good measure of citation behaviour type, in that it is an accepted measure of the distribution of inequality and can be applied to distributions of citation behaviour. To determine the Gini measure of inequality's applicability in the context of individual citation behaviours, it is necessary to see whether the assumptions Herdan used to choose the bilogarithmic form apply to this study.

The nature of linguistic studies of the form of text for which Herdan developed his measure meant that there was a potentially high degree of variation between text lengths or samples, particularly as a percentage of total text length. Drawing an analogy with the law of relative growth from biology, Herdan applies the bilogarithmic form of the type-token ratio to account for this potential growth. This study has collected individual authors' citations over the time period available within the ISI citation indexes, and so is comparing citation across authors within relatively constrained numbers of citations. Though there is certainly variation among the authors in regard to their extent of citations, this is not comparable to the magnitudes of variation in text length that Herdan was attempting to account for. Therefore, analysis of citation behaviour type can be conducted without the necessity of correcting for sample size in the manner proposed by Herdan. If the Gini is accepted as a measure of citation behaviour type, the correlation found with the citee-citation ratio would indicate that this is a preferable indicator of citation behaviour type than the bilogarithmic citee-citation ratio.

4.4.1. ASSUMPTIONS AND STATISTICAL DISCRIPTIVES FOR THE GINI VERSUS TYPE-TOKEN RELATION – APPROPRIATENESS FOR LINEAR REGRESSION ANALYSIS

The variables Gini and citee-citation ratio, and the unstandardised residuals (or errors) for the Gini versus citee-citation relation conform to the assumption of normality, in that their underlying distributions approach the normal distribution. All these elements have a skewness approaching zero, -0.638, 0.813 and -0.803 respectively, which are all within the acceptable range for the assumption of an underlying normal distribution (Tabachnick & Fidell, 1996, p.73). A test for normality was also performed, the One-Sample Kolmogorov-Smirnov Test, against a hypothesised normal distribution. The variables and the standardised residual significance values were found to be larger than 0.05, indicating that the observed distributions formed correspond to the normal distribution (Kinnear & Gray, 1999, p.143). However, it should be noted that the citee-citation ratio is only slightly above the 0.05 threshold at 0.059. When using the chi-square goodness of fit test for normality, it is also established that the chi-square significance for the Gini and the citee-citation ratio is below the critical value, and so it can be said that samples of these two variables come from a normally distributed population (Argyrous, 2000, pp.331-333).

The assumption of normality is related to the assumption of homoskedasticity (Tabachnick & Fidell, 1996, p.80). Homoskedasticity is the condition that the variance in prediction errors is more or less constant for the values of the independent variable, creating an even spread across and around the regression line; if this assumption is not met significance tests and confidence intervals are invalid (Lewis-Beck, 1980, p.28). Visual examination of the scatter of residuals confirms that their variance is even, validating that the assumption of homoskedasticity holds for this analysis (Argyrous, 2000, p.222). The assumption that the expected mean value of the error term is zero has also been met (Standard predicted value, mean = 0.000) (Lewis-Beck, 1980, p.26).

In order to test the assumption of linearity, that there is a straight line between the two variables, the standardised residuals from the regression analysis were plotted against the predicted values. The resulting spread of values is even, with slight accumulation centered around the zero residual value, indicating that the condition of linearity has been met (Tabachnick & Fidell, 1996, pp. 78, 136-137).

Statistically the assumptions of collinearity (or autocorrelation) are met by the Gini versus citee-citation relation. The Durbin-Watson statistic for autocorrelation approaches 2, being 2.082 and the collinearity tolerance statistic is 1, indicating that the dependant variable (Gini value) is largely independent of the independent variable (the ratio value). The collinearity conditioning index is well below 30 at 6.743 but the two variance proportions are above 0.50 at 0.98 each; for the collinearity assumption to be considered problematic the conditioning index would be both higher than 30 and the variance proportions would also be higher than 0.50 (Tabachnick & Fidell, 1996, p.87).

It should be noted however, that while the relation meets the statistical criteria for non-violation of the autocorrelation assumption, the measures used to generate the values are not entirely independent from one another. Both the Gini and the citee-citation ratio variables used within this analysis are calculated from the individual authors' citee and citation counts. The calculation of the citee-citation ratio is relatively straightforward as it is simply the individual authors citee count (N) divided by the total number of citations of that author, or to put it another way, the frequencies with which the citees were cited totalled, (T).

$$Ratio = \frac{N}{T} \quad (1)$$

The Gini value on the other hand is somewhat more sophisticated, in that it captures the inequality within the distribution of an individual author's citee and citation usage. It is calculated as follows:

$$Gini = \frac{N+1}{N} - \frac{2}{NT} \sum_{i=1}^N i \cdot x_i \quad (2)$$

where x_i is the number of citations for the citee of rank i . The expression of this formulation of the Gini is detailed in Section 4.4.2. below.

As can be seen, the measures are derived from the same underlying empirical counts, but are formulated in different ways, and so are measuring different aspects of an individual's citation behaviour. Thus the high correlation between the two measures can to some extent be equated with similar results obtained with very high collinearity, primarily indicated by an R^2 value over 0.9. Under such circumstances one or other of the variables would be eliminated as they are proved redundant through this result (Tabachnick & Fidell, 1996, p.84). However in this analysis the high correlation is presented to indicate the correspondence of the measures, that could not be determined by simply studying the formulae used to calculate them. Consequently this finding allows for the extension of the investigation of citation behaviours demonstrated by individuals, as shall be further elaborated throughout the following sections.

4.4.2. MATHEMATICAL EXAMINATION OF THE GINI VERSUS CITEE-CITATION RELATION

The formulation of the Gini is (Rousseau, 2000, p.7; Rousseau, 2001)

$$Gini = \frac{N+1}{N} - \frac{2}{N} \sum_{i=1}^N i \cdot a_i \quad (3)$$

where N is the total number of citees, i is citee rank and a_i is the number of citations for the citee of rank i divided by the total number of citations, T .

For the Gini to have a trivial relationship with the citee-citation ratio, then

$$Gini = 1 - const \times Ratio \quad (4)$$

and when plotting the Gini against the ratio, a straight line will result.

To explore this, the Gini can be rewritten as

$$Gini = \frac{N+1}{N} - \frac{2}{NT} \sum_{i=1}^N i \cdot x_i \quad (2)$$

where x_i is the number of citations for the citee of rank i . Note that is this the form of the equation (2) used above in Section 4.4.1. to demonstrate the comparative roles of N and T within the citee-citation ratio and Gini calculations. For large values of N ,

$$\frac{N+1}{N} \approx 1 \quad (5)$$

then for a trivial relationship to exist

$$\sum_{i=1}^N i \cdot x_i = \text{const} \times N^2 \quad (6)$$

This is certainly not true for any data set, as x_i is independent of N , therefore the relationship between the Gini and the citee-citation ratio is not trivial.

A further analysis can also be conducted to determine T as a function of N . If the Gini is accepted as having a linear relationship with the citee-citation ratio, an equation can be written in the form

$$\text{Ratio} = B - C \cdot \text{Gini} \quad (7)$$

where B and C are constants. B is simply determined by setting Gini equal to zero, where the citee-citation ratio is equal to one, based on the evenness criterion. Therefore

$$\text{Ratio} = 1 - C \cdot \text{Gini} \quad (8)$$

From the Consciousness data in this study, C is equal to approximately 1.18.

The above equation (8) can then be written as

$$\frac{N}{T} = 1 - C \left(\frac{N+1}{N} - \frac{2}{N} \sum_{i=1}^N i \cdot a_i \right) \quad (9)$$

or

$$\frac{N}{T} = 1 - C \left(\frac{N+1}{N} - \frac{2}{NT} \sum_{i=1}^N i \cdot x_i \right) \quad (10)$$

The relationship between N and T can then be determined. Multiplying both sides by N ,

$$\frac{N^2}{T} = N - C(N+1) - \frac{2C}{T} \sum_{i=1}^N i \cdot x_i \quad (11)$$

Then

$$\frac{1}{T} \left(N^2 - 2C \sum_{i=1}^N i \cdot x_i \right) = N - C(N+1) \quad (12)$$

or

$$T = \frac{\left(N^2 - 2C \sum_{i=1}^N i \cdot x_i \right)}{N - C(N+1)} \quad (13)$$

It is difficult to see from the above equation (13) whether T and N are related in the way described by Herdan (1960). This is principally due to the summation of rank multiplied by productivity, which is difficult to analyse.

4.4.3. EXAMINATION OF GINI VERSUS CITEE-CITATION RELATION AND VARIABLES THROUGH DATA VARIATION

When all citations from individual authors' citation distributions from this data set are included in the Gini versus citee-citation relation, the line of best fit intercepts the y-axis at between 0.8 and 0.85, and the x-axis at 1. Based on the evenness criterion of Gini, when Gini is equal to zero, the intercept at the x-axis can be determined theoretically. Generally, the Gini is equal to zero when even citation patterns are displayed. When single citations (or unicitations) are included as part of the author's citation pattern, the even N/T condition is where all citations are unicitations, therefore, $N=T$ and $N/T=1$. If unicitations are removed then the minimum citation level is 2, therefore the evenness criterion will be altered resulting in $N/T = 1/2 = 0.5$. It can be seen that if more citations are removed, the line will intersect the x-axis at $N/T = 1/(\text{min citation level})$.

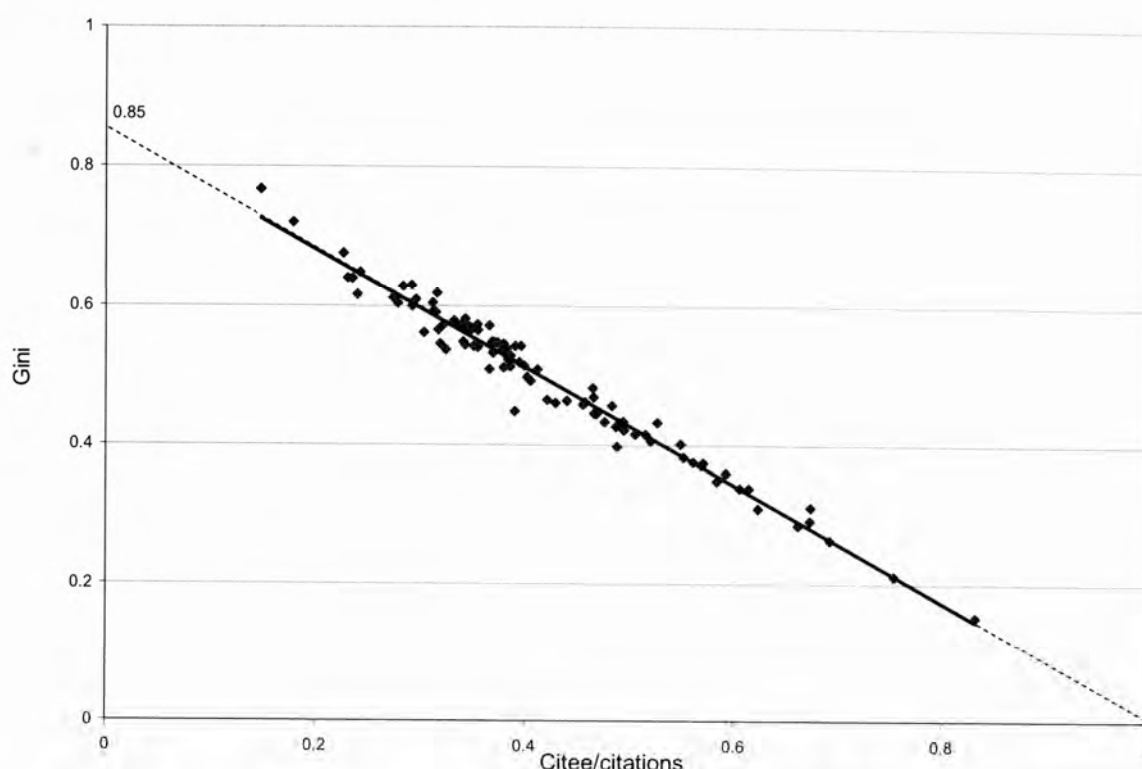


Figure 4.4 Placement of the theoretical slope (broken line) when all citations from authors are included

Figure 4.4 relates to the theoretical placement of the “average” line relating the Gini to the citee-citation ratio, although a linear fit to other data sets may result in the line being somewhat differently placed.

Placement of the intercept of the line with the y-axis is not as straightforward theoretically, as setting $N/T=0$ has no real meaning for citation studies. $N/T = 0$ can be achieved if T is significantly larger than N . This can be achieved artificially; however the resulting Gini can vary widely. Therefore the intercept is essentially determined by the nature of the data studied. For this data set, it is between 0.8 and 0.85.

An analysis of the fit of experimental data to the theoretical line was undertaken, by removing low level citations from the citation oeuvre of each author. The removing of lower level citations has the effect of increasing the level of scatter in the linear fit between the Gini and the citee-citation ratio. This was found to occur to some extent with unicitations removed, but to a greater extent when a five citation threshold was taken to be the minimum level of an individual author’s citation distribution.

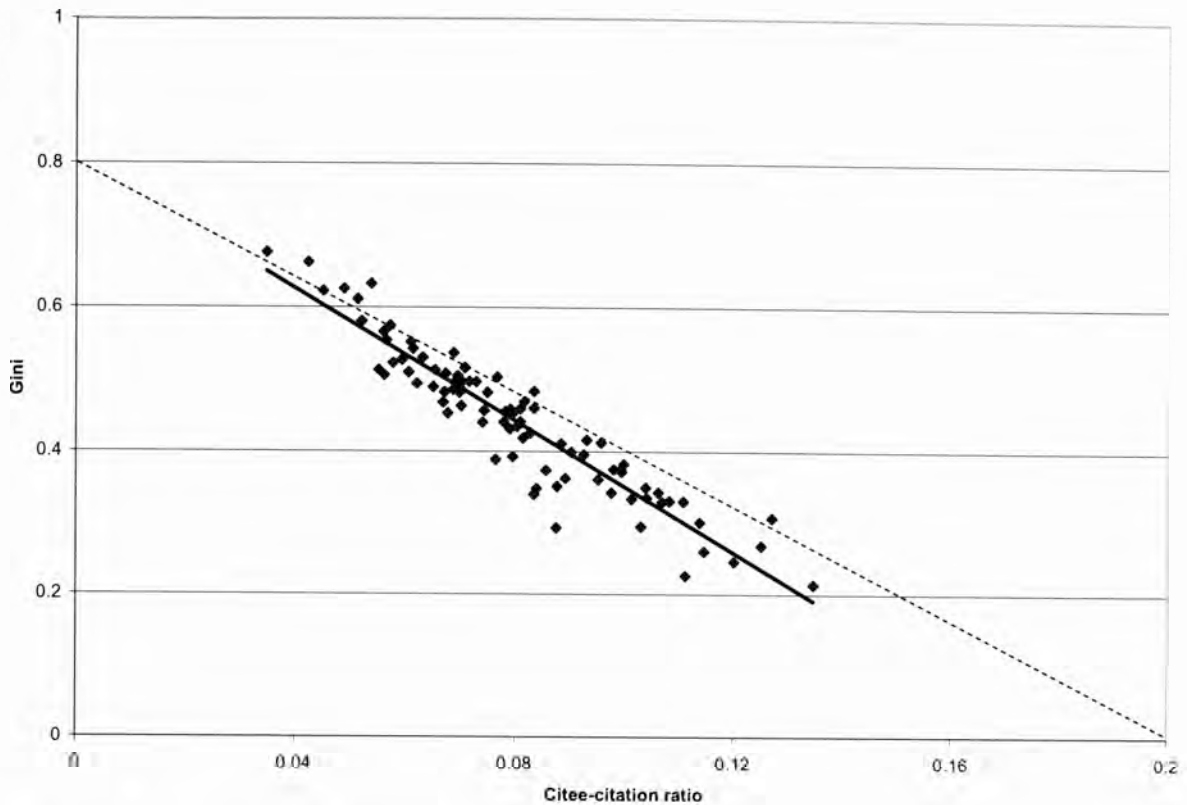


Figure 4.5 Placement of the theoretical slope (broken line) when citations from authors with a frequency below 5 are removed

As can be seen in Figure 4.5, scatter has been increased when the citation frequencies below five are removed from individual authors' citation distributions; this results in $R^2=0.88$ (compared with $R^2=0.97$ when complete citing identity distributions are included).

The slopes produced from the alternate data sets in Figure 4.4 and Figure 4.5 (solid lines) additionally have different positions in relation the theoretical lines (broken lines) in the respective figures. So it can be seen that the Gini value begins to depart from the predicted value in the alternate data sets for a given citee-citation ratio value.

A discussion of the nature of the changed citation profile is presented for the departure of Gini values from theoretical values. Figure 4.6 shows two potential citation patterns for an individual author. Here the citee-citation ratio is the same for both curves, and the total number of citations and total number of citees is also the same, as they meet at the same end point in the graph. The Gini coefficient of the blue curve will be lower than that of the black curve.

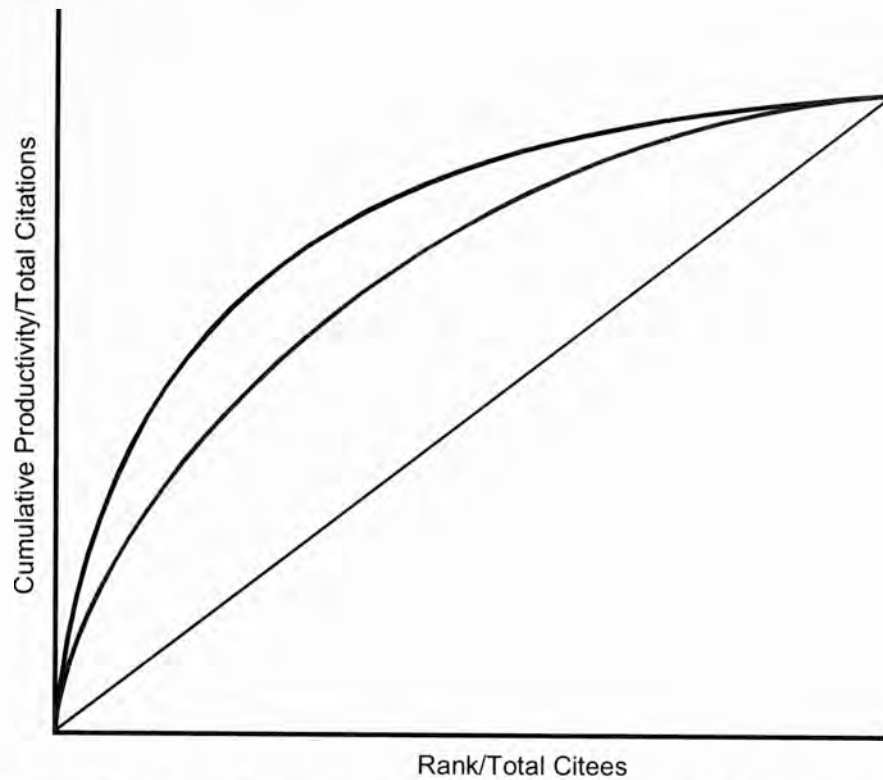


Figure 4.6 Schematic of an author's citation distribution with equal citee-citation ratio and varied Gini values

An explanation of reduction in the Gini value is made with reference to Figure 4.7, which shows level of citation versus the number of citees with the corresponding level of citation, for the two cases identified in Figure 4.6. The total number of citees (N) and citations (T) are equal. Furthermore as the product of N and T is also equal, the area under each curve in Figure 4.7 must also be equal.

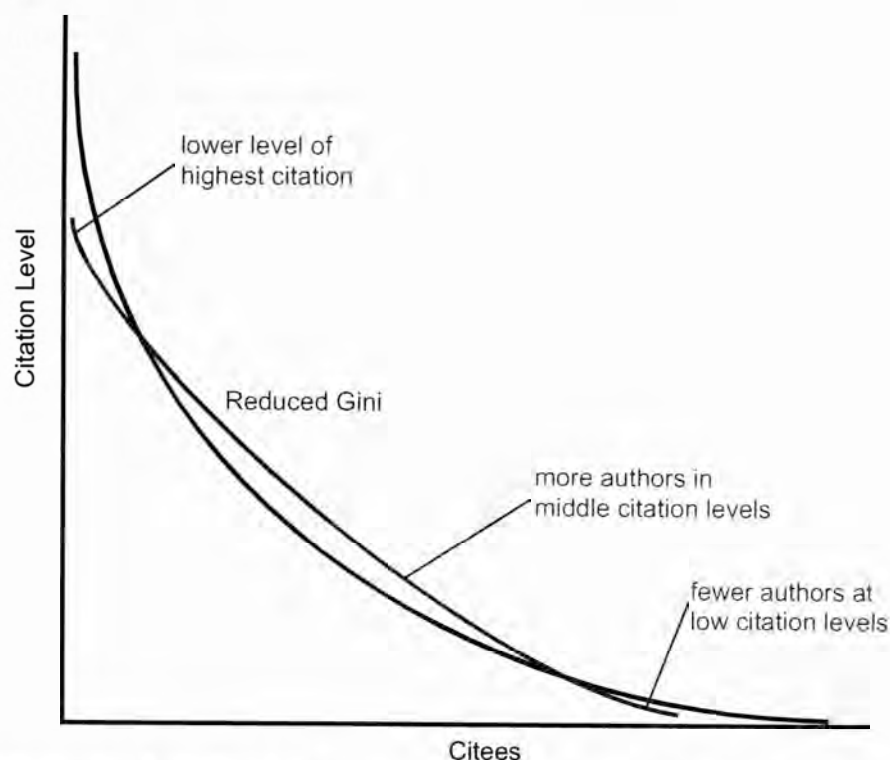


Figure 4.7 Schematic of citation level versus citees for citation distributions identified in Figure 4.6.

In this example (Figure 4.7), in the case of the reduced Gini (blue line), there are less citees at the lower end of citation level frequency, and also the highest citation level is lower. This leads to the curve with the lower Gini (blue line) being above the higher Gini curve (black line) for the middle points within the citee's citation level range. For example, there may be more authors with 10-30 citations, but no-one with 1200 citations.

Returning to Figure 4.5, the higher Gini (black curve) in Figure 4.6 and Figure 4.7 may represent data that sits on the theoretical (broken) line, whereas the reduced Gini (blue curve) may be seen to correspond with data that fits the actual trend line for the Gini versus citee-citation ratio relation. In this instance it is expected that the actual trend line would correspond with the theoretical line for this case if the authors' citation distributions better conform to the conditions of the higher Gini. Currently the distribution that defines the higher Gini (black line) is unknown.

4.4.4. IMPLICATIONS OF GINI VERSUS CITEE-CITATION RATIO FINDINGS

The above analysis was undertaken to find a simple explanation for the tight relationship between the Gini and citee-citation ratio. A comprehensive analysis of the mathematical underpinnings of the relationship, as well as further exploration of modified data, did not reveal a simple explanation. In fact, the establishment of a theoretical fit line, and the ability to modify data to cause a departure from that

theoretical line, indicates that the relationship between the Gini and citee-citation ratio is not one that can be explained by the existence of common underlying variables.

The implication from this finding is that a regularity is present in authors' citation oeuvres, that gives rise to the tight and predictable relationship between the Gini and cite-citation ratio. This regularity requires further exploration, specifically whether this regularity is identified by one of the existing informetric "laws".

4.5. GINI VERSUS CITEE-CITATION RELATION AND THE INFORMETRIC "LAWS"

4.5.1. ZIPF'S LAW

Zipf's Law expressed in the rank-frequency form is (Wilson, 1999, p.177, Equation [4c])

$$x_i = \frac{D}{i} \quad (14)$$

where i is the rank, x_i is productivity and D is the highest productivity. In this case the Gini as described in equation (2) is (Egghe & Rousseau, 2003)

$$G = \frac{N+1}{N} - \frac{2}{NT} \sum_{i=1}^N D \quad (15)$$

Then

$$G = \frac{N+1}{N} - \frac{2ND}{NT} \quad (16)$$

or

$$G = \frac{N+1}{N} - \frac{N}{T} \times \frac{2D}{N} \quad (17)$$

Here the Gini would have a linear relation with the citee-citation ratio (N/T) if the factor $2D/N$ was constant. However, for the data set of consciousness researchers, the highest D/N for an author is fifteen times as large as the lowest D/N . Therefore Zipf's Law cannot be used to explain the linear relation between the Gini and the citee-citation ratio.

Explanation of the Gini versus citee-citation ratio linear relation using other informetric "laws" is complicated by the sum of rank times frequency contained within the Gini equation, therefore if an informetric law can be used to explain the Gini versus citee-citation ratio its proof will be complicated.

4.5.2. POWER LAW AND LOGORITHMIC CITATION DISTRIBUTIONS

Individual author's citation distributions were studied for their level of fit to a power law curve. The normalised cumulative productivity versus normalised citee rank was plotted and a power law curve fitted using the standard function in Excel. The R^2 coefficient resulting from applying this curve fit was then plotted against the Gini coefficient.

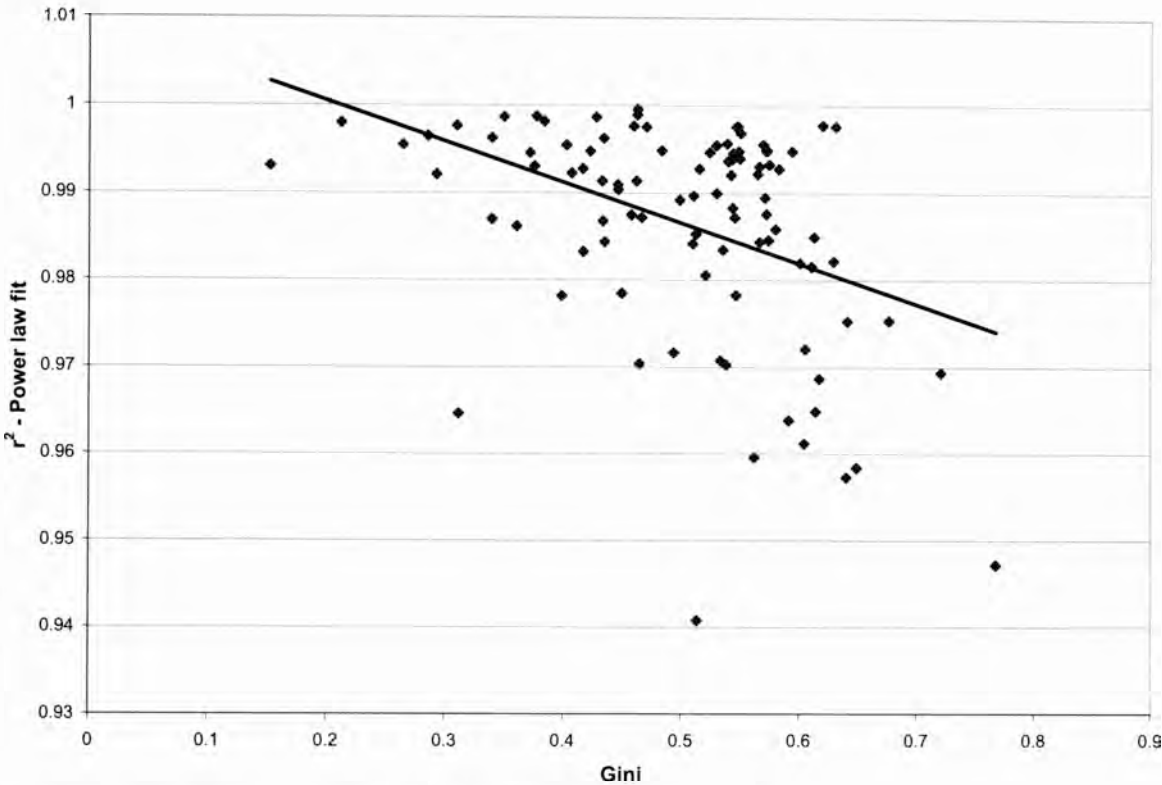


Figure 4.8 Power law fit R^2 coefficient versus Gini coefficient

The results of this can be seen in Figure 4.8. A linear trendline is also plotted which itself has a R^2 correlation of 0.17, identifying that a fit to power has no correlation with position on the Gini versus cite-citation ratio line. This is an important distinction to make, as none of the authors can be classified as having a “poor” fit to the power law curve. Notable in this figure is a high number of authors with a very good fit to a power law curve; however a small but significant number of authors do not fit as well. At this point another form of curve fitting was attempted for all of the authors' normalised cumulative productivity versus normalised citee ranks, to examine whether alternative forms of curve fitting would better suit some authors. A logarithmic trendline was attempted on each author's normalised cumulative productivity versus normalised citee rank and the R^2 coefficient plotted against the Gini, as can be seen in Figure 4.9.

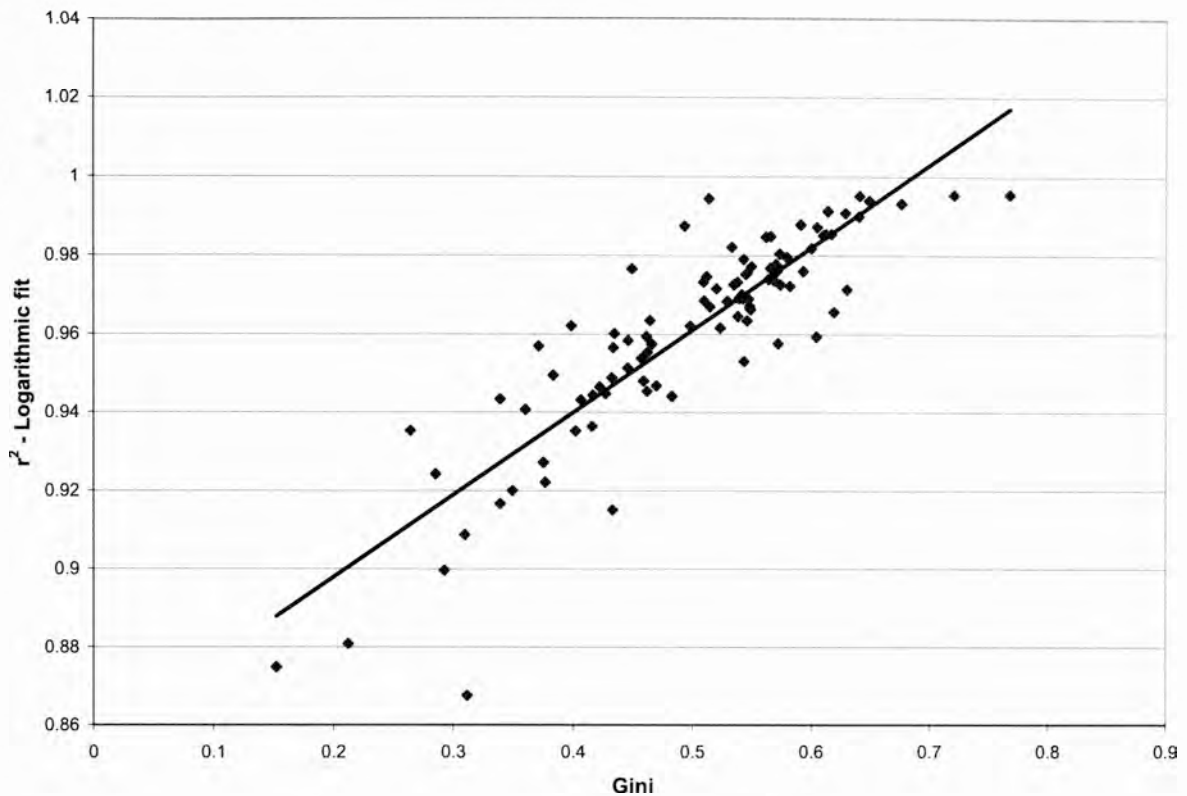


Figure 4.9 Logarithmic fit R^2 coefficient versus Gini coefficient.

The plotted linear trendline in this case (Figure 4.9) resulted in an R^2 of 0.77. This form of analysis therefore appears to be more discriminating between authors than using the power law fit to the authors' normalised cumulative productivity versus normalised citee rank. Additionally there were relatively few authors with very high logarithmic correlation and those authors in general also have a high value of Gini coefficient. In order to see the value of logarithmic trendline fitting, authors with a better logarithmic fit than power law fit were identified and the Gini versus citee-citation ratio replotted.

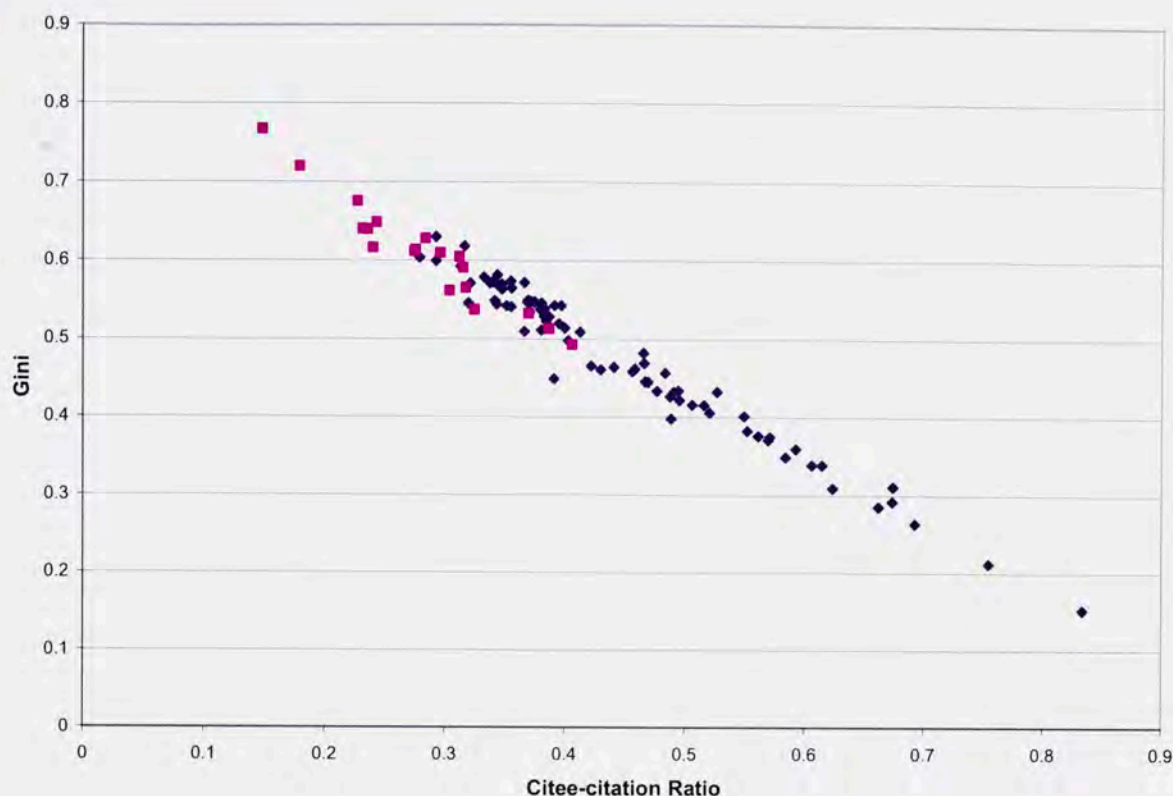


Figure 4.10 The Gini versus citee-citation ratio relation with (■) logarithmic fit better than power law fit, and (◆) power law fit better than logarithmic fit.

A high level of separation is apparent (Figure 4.10), where authors with higher logarithmic fit are placed toward the upper left corner of the Gini versus citee-citation ratio graph. There is a group of three authors with the highest citee-citation ratio of the group of authors with higher logarithmic fits. It is notable that two of these authors (MARI JIBU and GIUSEPPE MORUZZI) have low total citation counts, each at 207, although the third author (PAUL BLOOM) has a notably higher count, at 1247. The average total citation count for authors with a higher logarithmic fit is 4016 compared to 2069 average total citations for authors with a higher power law fit. Therefore as a group, authors with a better logarithmic fit tend to have a higher number of citations. However, when comparing the authors with the highest citation counts, the first and second ranked authors (GERALD EDELMAN and JEFFERY GRAY) have better power law fits while the third and fourth ranked authors (WOLF SINGER and DANIEL SCHACTER) have better logarithmic fits.

The fact that some authors have a better fit to logarithmic distribution raises the question as to the applicability of the Bradford distribution. An author with a high logarithmic correlation (THOMAS NATSOULAS) and an author with a high power law correlation (FRANCIS CRICK) were compared on the basis of their separate individual citee-citation plots.

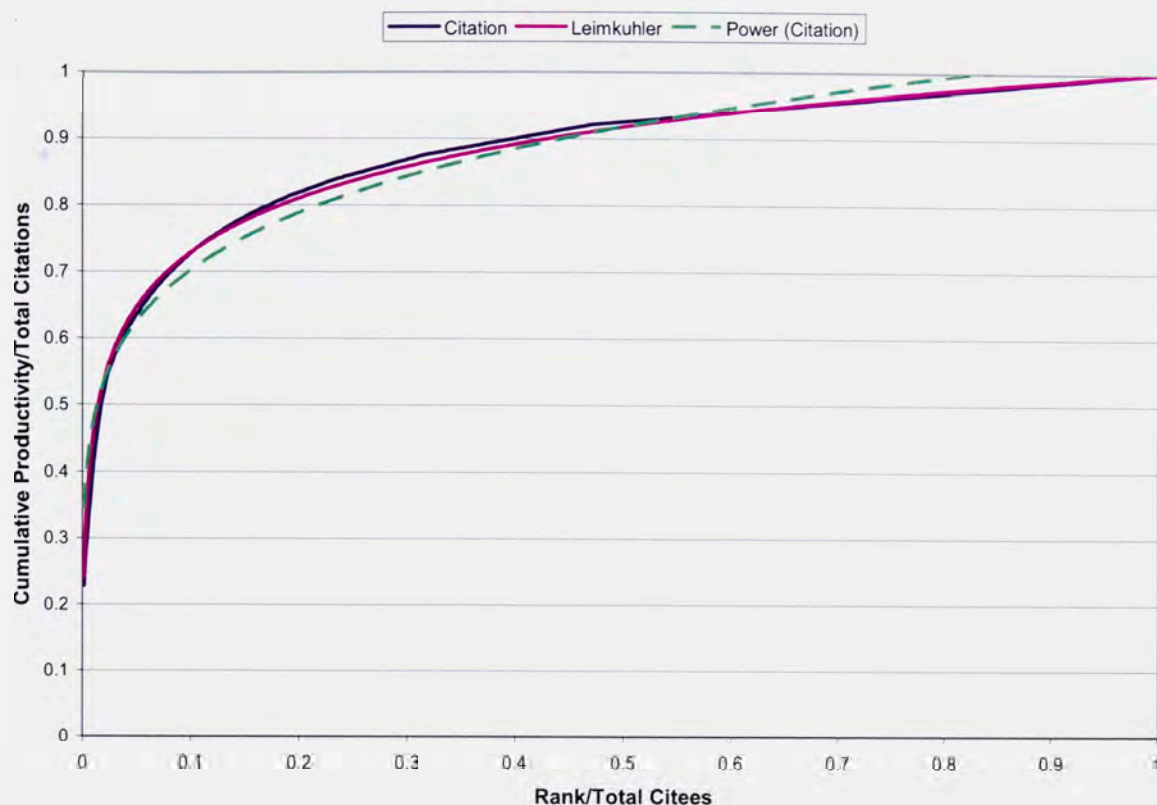


Figure 4.11 THOMAS NATSOULAS' individual citee-citation plot

Figure 4.11 shows NATSOULAS' citation distribution, a power law fit to the citation distribution and a Leimkuhler form of Bradford's distribution (Leimkuhler, 1970), calculated by (Wilson, 1999, p.176, Equation [5b])

$$X(x) = \frac{\ln(1 + Dx)}{\ln(1 + D)} \quad (18)$$

where x is rank divided by total citees, $X(x)$ cumulative productivity divided by total citations and D is a constant. Forms of the Bradford distribution were expected to correlate well with data that had a good logarithmic fit. Leimkuhler's form of the Bradford distribution was found to fit particularly well with NATSOULAS' citation distribution.

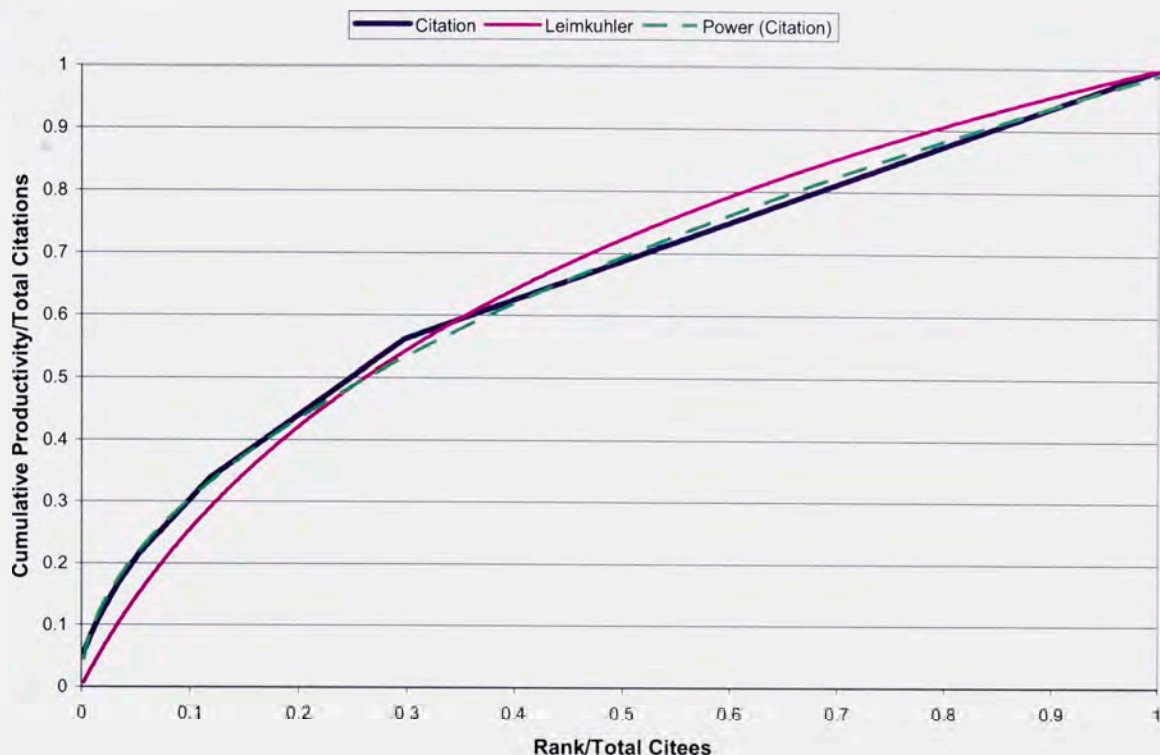


Figure 4.12 FRANCIS CRICK's individual citee-citation plot

For comparison, Figure 4.12 shows CRICK's citation distribution, a power law fit to the citation distribution and the Leimkuhler form of Bradford's distribution. CRICK's citation distribution better approximates a power law fit than the Leimkuhler form of Bradford's distribution. Therefore a distribution that has a power law as its basis may be more suitable to describe CRICK's citation distribution than a form of Bradford's distribution.

In summary, these results demonstrate that it is not likely that all of the authors' citation distributions will be described by any single distribution based either on power laws or logarithmic distributions, which are the basis of some of the current informetric "laws". Therefore, the relation identified between the Gini coefficient and citee-citation ratio cannot at this stage be attributed to one of these "laws". However what may result is that the relation between the Gini coefficient and citee-citation ratio can be used to discriminate effectively between authors, based in their citee-citation distribution, and further exploration of author differentiation is therefore warranted.

4.6. APPLICATION OF THE GINI VERSUS CITEE-CITATION RATIO RELATION

The identification of a regular relation between the Gini and citee-citation ratio in the boundary specialty of Consciousness was outlined in the previous sections. Noting the

diversity of the boundary specialty in question, the identification of the relation is thought to be relatively independent of the data set on which it is based. The reapplication of the relation to additional data sets is not within the scope of this work, and therefore full verification of the applicability of the relation cannot be assumed at this stage. However, a preliminary exploration of the use of this relation in understanding of authors can be undertaken.

There are several points to note regarding the underlying consistency of the relation. Firstly, the high level of consistency identified within the Consciousness data set shows that there is a level of consistency within an author's citation oeuvre; as identified in Section 5.4, this cannot yet be attributed to any identified regularity or "law" described in the literature. Secondly, the finding of this tight relation between Gini and citee-citation ratio has given some weight to White's use of citee-citation ratio as a distinctive measure of individuality. It is therefore possible (within the Consciousness authors data set) to view Gini and citee-citation ratio as unambiguously related. The Gini will continue to be used throughout the remainder of this work, although it is a more difficult number to calculate. This is because the use of a concentration measure (i.e. level of departure from even citation rates) is more readily understandable as a discerning descriptor of a citation oeuvre than a ratio of total citees to total citations.

The most significant point of discussion is the possibility of using the identified relation as the basis for differentiating authors within a data set. Each author in the Consciousness data set can be placed on the same baseline, i.e. each has a citation oeuvre that has a set relationship between Gini and citee-citation ratio. Furthermore each author can be differentiated from other authors within the data set by use of a single variable (Gini).

The question of citing identity is raised in relation to this final point. A similar concentration coefficient, such as measured by the Gini coefficient, was illustrated in Section 4.3 with a discussion of COWEY and HOBSON. These researchers presented similar citee-citation ratios and identical Gini values; however were distinctly different in terms of some traditional measures of citation, namely highest cited author counts and uncitation counts. On the surface this would indicate that the tight relation between Gini and Citee-citation ratio is not sensitive to number of uncitations or highest cited author count. However it is also possible that the Gini coefficient, as an aggregate number, is not sensitive to these individual factors *taken in isolation*. The extension of this proposition is that the Gini, as an aggregate number, is a usefully

consistent measure for discerning authors. Neither this proposition, nor the proposition of lack of sensitivity, can be adequately tested within the scope of this work.

In either circumstance, the relation can be exploited to discern between authors' citation oeuvres in a consistent manner. This potentially allows definition of citation types or identities based purely on citation data, without the requirement for interpretation by an "informed researcher". The creation of data-generated citing identity types is the focus of the following section.

4.7. ESTABLISHING CITING IDENTITY BEHAVIOUR TYPES

Cluster analysis was used to establish a citation behaviour typology, according to the similarity across the combination of the variables Gini and citee-citation ratio. This method does not entirely eliminate arbitrary groupings, in the sense that the choice of the clustering algorithm applied directly determines the resulting typology, and can impose structure rather than simply uncover natural structure within the data (Everitt, Landau, & Leese, 2001, p.8). However the use of cluster analysis does allow for a systematic identification of types (Anderberg, 1973, p.4). Cluster analysis also provides a method for establishing a typology that is able to be replicated. The use of somewhat arbitrary groupings for the purpose of examination of citation behaviour types according to the two variables is not a significant limitation within the scope of the current work. These types will be examined in relation to biographical and bibliographic characteristics of the individual authors in Chapter 5, in order to determine whether placement according to Gini and the citee-citation ratio values, either individually or combined (as there is a demonstrated close correlation between the two variables), varies with regard to an individual authors identifiable social or cognitive characteristics.

Accordingly for this analysis the primary consideration for validation of the alternate grouping solutions for a typology is their adequacy in representing the relationships between the variables, Gini and the citee-citation ratio. This is referred to as external validity (Everitt et al., 2001). Various solutions, in terms of both alternative cluster methods and the resultant appropriate number of groupings have been considered and compared to the relation between Gini and citee-citation ratio (hereafter referred to as the relation). Centroid, single linkage or nearest neighbour, k-means and Ward's methods have been examined. All clustering methods have a number of limitations in their usage and so therefore 'the main problem in practice is that no particular clustering method can be recommended, since methods with favourable mathematical

properties (such as single linkage) often do not seem to produce interpretable results empirically' (Everitt et al., 2001, p.89).

In the development of an appropriate typology an approach has been taken to attempt to compensate, where possible, across limitations between methods to arrive at an adequate representation of clusters for the purpose of representing the relation. Single linkage or nearest neighbour clustering has the limitation of tending to cause 'chaining', where singleton or very small clusters with little structure are produced; however the advantage to this method is that it can highlight outliers (Everitt et al., 2001, p.64). The single linkage solution from clustering of the Gini and citee-citation ratio did produce a dendrogram with little structure; yet it also highlighted a number of singletons that have consistently appeared as separate clusters within other methods, and also when comparing different numbers of cluster levels. These results were then used to modify the definition of the number of clusters within other methods. Single linkage can either be applied as an agglomerative or as a divisive method. Agglomerative methods, such as Centroid and Ward's method, overcome the problem of 'chaining' by building the clusters into one whole, rather than by splitting the individuals off from the whole. However, the Ward's and the Centroid methods have the opposite tendency: to produce clusters that are spherical in shape and of the same size, having the effect of potentially obscuring appropriate divisions in the clustering solution (Everitt et al., 2001, p.62).

For this analysis the original Ward's method solution identified five clusters. These incorporated the outliers that were identified in the single linkage solution into larger clusters; though when a higher number of clusters was specified, at the level of seven clusters, the outliers became identified within the Ward's method solution. At the level of seven clusters, the Ward's method and the Centroid method produced the same partition structures and grouped together the same individuals. The k-means procedure also resulted in a very similar structure and groupings, the only exception being that IAN MARSHALL was grouped as a singleton whereas in the Ward's and Centroid methods he was grouped with RICHARD WATT. Distances in the relation line can support either grouping. However, given that both are spaced at comparable distances from the nearest group, placing them together can be justified, though the potential differences this may reflect should be noted.

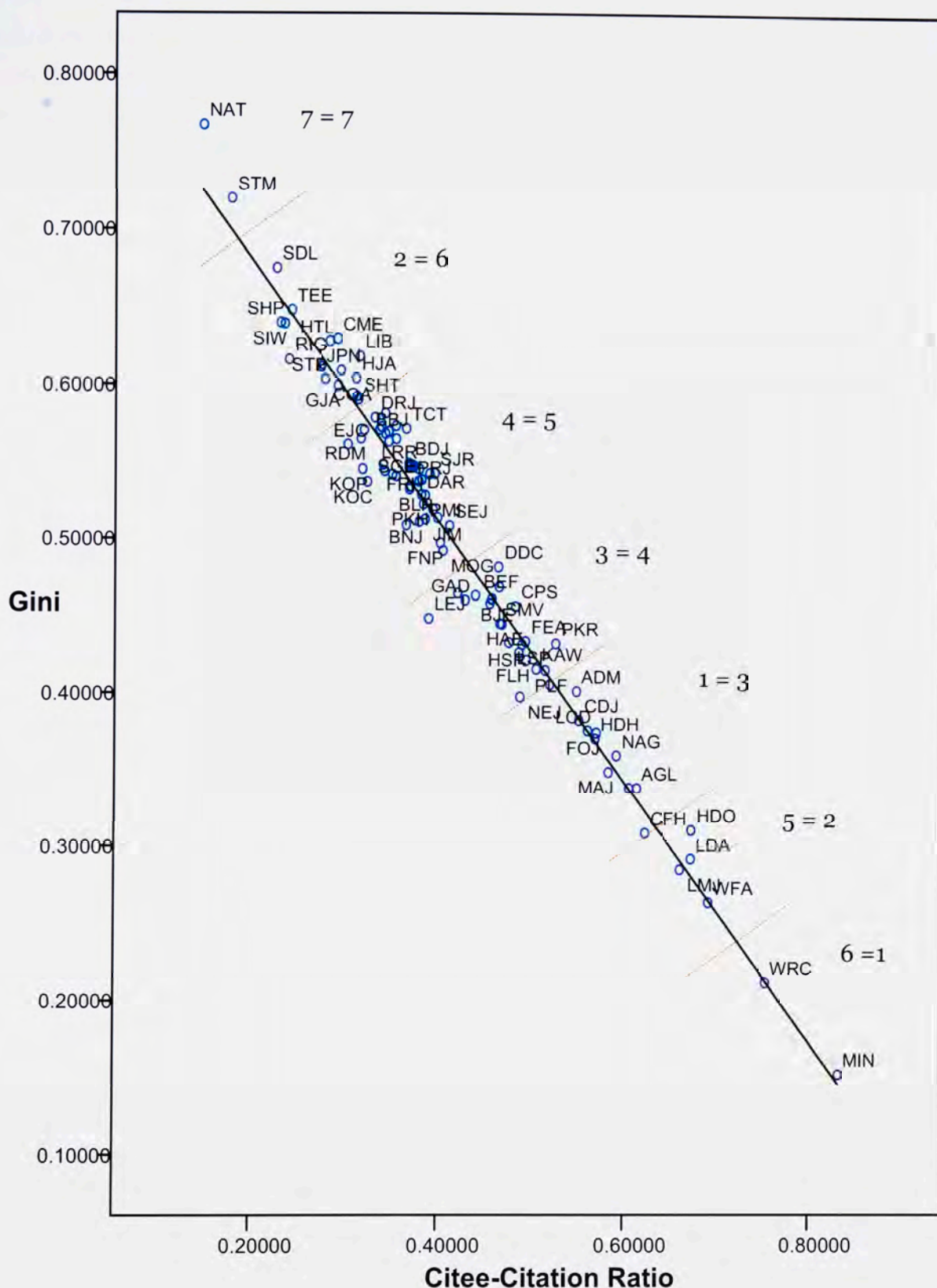


Figure 4.13 Ward's and Centroid cluster analysis compared with the Gini and Citee-citation relation

Comparison of the k-means, Ward's and the Centroid methods and the finding of a large degree of correspondence between them at the level of seven clusters adds confirmation for the usage of this typology for the examination of citation behaviour

types based on the relation.⁴⁰ Figure 4.13 demonstrates the Ward’s and Centroid cluster solutions superimposed on the Gini and citee-citation ratio relation. When comparison with the citee-citation ratio alone is considered, the resulting clusters approximately correspond to the increase or decrease in the ratio value, with the exception of JOSEPH LEVINE and DOV SAGI, who have comparatively lower Gini values than the other members within the particular cluster and so have been considered by the clustering procedures as having a closer proximity to, and so have been placed within, the next nearest cluster.

4.8. CITING IDENTITY BEHAVIOUR TYPES

An empirical analysis of the comprehensive publication records of ninety-seven scientists in the field of consciousness research, comprised of disparate disciplinary backgrounds, has shown a strong regularity in citing behavior. There is a high correlation between the ratio of cited authors to citations (Citee-Citation ratio) and the Gini index, a measure of dispersion of citing. This regularity has been further analyzed using cluster analysis to establish citing identity types.

The cluster analysis methodology has been undertaken to establish type definitions without qualitative interpretation. Ideally this provides a data driven segmentation of the candidate authors citing identities. In practice determining the success of the methodology requires testing against bibliographic variables; this further examination of authors’ characteristics is presented in Chapter 5.

However without this detailed examination of the methodology, some preliminary characteristics of the citing identity types can be presented. The cluster analysis preformed has enabled a distinction of seven types of citing behavior within the Consciousness author data set. The significance of the mid-range citing types is difficult to determine without further examination and determination of a theoretically informed method of delineation. In the cluster solution presented types 2 to 6 form a continuum between the two extremities, this delineation imposed only methodologically at this stage through cluster analysis to allow for further examination. However, the two extremities (types 1 and 7) can be described in terms of the citing behaviour determined by the relation.

Table 4-1 Comparison of citing identity types represented in the relation

⁴⁰ A table for comparison of the results for this analysis can be found in Appendix C, along with the dendogram comprising the final Ward’s method seven cluster solution.

Type	Number of cites	Number of authors cited	Citee-Citation Ratio	Gini index
Type 1	232	175	0.75	0.21
Type 4	2737	1154	0.42	0.47
Type 7	4852	721	0.15	0.77

As Table 4-1 shows Type 1 is a scientist who seldom re-cites authors, and therefore refers to most of the authors he or she cites only once, and ‘uses’ all cited authors with a similar frequency. Type 7 is a scientist who frequently re-cites authors and refers to authors unevenly, e.g. cites few authors very often, while others are cited much less frequently.

The importance of using a replicable method has been stressed as these types will form the basis of the following analysis (in Chapter 5) as the variations demonstrated by this finding should be explained by a theory of citation, which links variations of causes to variations in patterns of citations. Chapter 5 extends analysis of the citing types through the introduction of methodologies examining potentially relevant variables in the establishment of a communicated scholarly identity as developed in Chapter 3. For the purposes of this study examination of scholarly identity has been limited to the expression of an individual’s formally communicated identity through the introduction of the definition of research personas in Section 5.2.

4.9. CONCLUSION

An analysis of citee-citation concentration has been undertaken in the boundary specialty of Consciousness. Comparison of the authors’ citation oeuvres was examined, using the Gini coefficient and ratio of total citees to total cites to perform analysis. A high level of consistency was found for Consciousness authors between citee-citation ratio, and the Gini coefficient. The result is a linear relation between Gini and citee-citation ratio, with a R² correlation coefficient of 0.972. An exploration of theory and underlying data was undertaken, with the result that a trivial explanation could not be found for the high level of consistency. Additionally, the consistency could not be attributed to informetric “laws”.

The finding of the above relation was discussed primarily in its implication for generation of citing identity types, as the relation potentially provides both a consistent baseline for analysis and simple discriminator between authors. Subsequently, citation types were generated using Ward's and Centroid cluster analysis, with seven separate types generated. Analysis of Consciousness authors in Chapter 5 will include utilisation of the citing identity types developed from the relation and its variables, to examine citation behaviour with regard to researchers within their representational space attributes in the context of the boundary specialty of Consciousness studies.

5. CITING IDENTITY TYPE ANALYSIS, RESEARCH PERSONAS AND REPRESENTATIONAL SPACE

5.1. INTRODUCTION

The value of findings in Chapter 4 was the generation of a simple means of differentiating authors based on their citation profile. Based on the work presented in this chapter, a new data-driven method of analysing the individual compared to others in the field has been generated. The result is highly regular, despite the potentially disparate nature of the field chosen. Subsequently, citing types were generated using cluster analysis, to provide a fully data-driven method of analysing author identity through citation type.

The present chapter will contain an analysis of this new means of author characterisation, through comparing this and other methods of analysis. Preliminary analysis is initially conducted using accepted bibliometric techniques to provide placement of authors within the boundary specialty of Consciousness. Subsequent biographically informed, bibliographic and content analyses follow, utilising the findings from Chapter 4. Finally, the degree to which there is commonality between the analysis methods is presented, and a discussion of the discovery of a research identity through the different analysis techniques pursued.

5.2. RESEARCH PERSONAS AND BOUNDARY SPECIALTY REPRESENTATIONAL SPACE

The review presented in Chapter 3 highlighted consideration of the representational work practices undertaken by individual researchers in developing their research identities, and that these practices are further undertaken within the context of their knowledge production units. Researchers place themselves within their intellectual fields through developing reputations in part based on communication of the task outcomes of their scientific work practices as well as communicating their professional identities through the products of their scientific work, which further communicates the outcomes of their problem choice and its situation in terms of their membership and participation within intellectual fields. Formal communication is an outcome of these processes and is represented by the products of the formal communication system, publications.

In order to analytically distinguish the representational aspects from the more strategic and intentional aspects of an individual's research identity and communicative work

practices, the conceptualisation of a research persona is introduced. In this context a research persona is the characterisation of an individual as they represent themselves, in negotiation with their communities, via their publications, and so is a representational outcome of the individual's communicative relationships with their research networks. A research persona represents the positioning of a researcher in the communicative space of broader scientific networks. Positioning and communicative processes of an individual involve the researcher's use of respective knowledge bases, which contributes to the definition of an individual's research persona.

Unlike Lievrouw's (1996) scholarly identities and research narratives, research personas are specifically an outcome of the scientific and communicative work practices of an individual researcher as represented via their accumulated publications. Persona has the multiple connotations of being possibly, though not necessarily, self-establishing, advertently or inadvertently in some way representative of characteristics of the individual, as well as incorporating consideration of how the individuals' characteristics are perceived by others.

Research personas fundamentally draw on White's conceptualisations of citation identities (White, 2000, 2001b) and CAMEO's (White, 2001a), though are further distinguished as they are specifically communication and representational outcomes of scientific work practices. These practices involve the positioning of the researcher and their development of a representational identity through which they are perceived and interrelate to others in their knowledge production settings. Citation identities refer to the citing patterns of usage of other authors' work formed during a writing 'career'. White's (2001a) conceptualisation of CAMEO's is very similar to the definition of research persona as presented here, as both are reliant on the representational outcomes of an individual's communication practices within a formal publication environment. However White's CAMEO's involve an automated characterisation of the researcher, which further involves a number of system limitations that this conceptualisation of research personas as operationalised will seek to minimise.

The research persona conceptualisation maintains the central importance of the role of the individual and their work practices in knowledge production, and acknowledges that communication and community membership is a crucial component of the scientific work practices of individuals. However, this conceptualisation also recognises that there is an inherent limitation of representations to capture and therefore directly examine underlying scientific work practices and processes of which they are specifically, in their final published state, the behavioural outcomes.

5.3. METHODOLOGY

5.3.1. COLLECTION OF BIBLIOGRAPHIC AUTHOR DATA

In all cases, except for the author co-citation matrix⁴¹, the bibliographic records from the ISI citation indexes that were used for developing citing identities, as outlined in Section 4.2.5, were used to extract further bibliographic variables for the analyses presented below. This assisted in ensuring that the bibliographic data being used was the target authors', therefore to some extent reducing the homonyms and allonyms problem related to all bibliographical dataset establishment⁴², and that the bibliographic variables were referring to the same publication set as developed for the citing identities for comparative purposes.

Bibliographic variable data was extracted from the original Dialog records using a purpose written program to also extract cited reference data.⁴³ The variables extracted by this program include authors listed, journal titles, document types and publication year for each record. Once extracted, variables were further organised and analysed using Excel and SPSS. Further verification and data standardisation was needed to, as much as possible, ensure consistency and correct homonym collation. For example, co-author names and journal titles were further verified with biographical and website information for authors and a combination of journal publisher details available via publisher websites, and Ulrich's international periodicals directory for journals.

5.3.2. RESEARCH PERSONAS OF PARTICIPANTS IN REPRESENTATIONAL SPACE: CONSCIOUSNESS BOUNDARY SPECIALTY AND CITING IDENTITY TYPES

5.3.2.1. INTELLECTUAL PLACEMENT OF PARTICIPANTS BY OTHERS - AUTHOR CO-CITATION ANALYSIS

As discussed in Chapter 3, individual authors and scientific researchers must be positioned within an intellectual space in order for a research career to be possible. Author co-citation is a methodology that achieves a contextualised representation of an author through the use of their work by others within their field. Author co-citation matrices and analysis represent the use of an author's work by others in the field,

⁴¹ See Section 4.2.3

⁴² Allonyms is a terms coined by White to refer to variations of an individual author's name, whereas homonyms refer to different authors with the same name - see (White, 2001b, p.91) for a detailed discussion

⁴³ See Section 4.2.5 4.2.5and Appendix B.1.

through counts of the number of times a cited author is cited with another cited author by a citing author. It is assumed that close use of two authors work by other authors represents a similarity in research focus or intellectual positioning. An analysis of the intellectual positioning of individual authors was performed in order to examine the question of whether citing identity types varied in any patterned way across the boundary specialty of Consciousness.

Participant authors for the author co-citation analysis were identified and a matrix developed during the Consciousness boundary specialty identification process as outlined in Section 4.2.3.⁴⁴ With the exception of 3 of the 100 authors identified (Sigmund Freud, William James and Wilder Penfield) the authors whose citing identities form the basis of this study are also the authors whose co-citation relationships are represented in this author co-citation matrix. In this regard the author co-citation matrix forms an external representation of use relationships of the authors under study. Co-citations are formed by the combination of each author in the set with each other author in the set in order to produce the matrix, so the author co-citation representation is not how the authors under study identify themselves, but rather how they have been placed in the field by others that use a combination of their work with any other author in the set. This is a partial capturing of the concept White refers to as an author's 'image' and 'intellectual structure', indicating the placement of an individual's work in relation to others in the field.

The similarity that is therefore addressed is the frequency of co-occurrence of an author's work with another author's work, in the work of any author that has combined them. The authors' semantic similarity is only assured to the extent that it can be assumed that their work has been used concurrently within another contribution, and that this implies a common usage relationship that is communicatively meaningful between the two authors. The resulting similarities and interpretations of these similarities are then examined in light of the individual authors' citation identities to determine the extent to which authors that have been placed in close proximity to one another in terms of use, also have similar citing identity distributions.

5.3.2.1.1. VISUALISATION OF THE AUTHOR CO-CITATION MATRIX

⁴⁴ Further detail on the determination of the Consciousness author set is provided in methodology Section 4.2.3 – a notable limitation for this analysis is the use of Dialog to collect co-citation data, Dialog access to the ISI citation indexes only allows access to the first author listed in a cited work.

Multidimensional Scaling (MDS) techniques were applied to the ‘raw’ author co-citation matrix. A PROXSCAL (SPSS)⁴⁵ solution was used with no further proximity measures calculated. Leydesdorff (2005 preprint) recommends this procedure as co-citation matrices are already similarity matrices by their nature, as they contain collected counts based on the similarity of use by citing authors. As co-citation is already a proximity matrix, the principle remains that higher the co-citation (proximity), the more similar the units are, and the closer the authors will be placed in the map. In order to reduce stress (and therefore improve the fit between solution and matrix) in PROXSCAL, the data is treated as ordinal and “untie the observations” is selected (Leydesdorff & Vaughan, 2005 preprint, p.22).

5.3.2.1.2. AUTHOR CO-CITATION MATRIX CLUSTER ANALYSIS

In order to create a systematic means of comparison a cluster analysis was also completed on the author co-citation matrix (Anderberg, 1973, p.4). Cluster analysis also has the advantage of providing a systematic and replicable typology more directly comparable to the citing identity types.⁴⁶

A number of solutions were examined and compared for use. The Average linkage (within groups) cosine solution was preferred as it avoided the problems of either not discriminating adequately between clusters (other solutions produced over 50% of authors within one group) or significant amounts of chaining.⁴⁷ The MDS and the cluster analysis were then used to examine citing identity types and representational intellectual placement within the Consciousness participant set.

5.3.2.2. COLLABORATIVE REPRESENTATIONAL SPACES - CO-AUTHOR ANALYSES

Co-authors necessarily share a proportion of their publications, therefore they will also share a commensurate portion of their cites and their citing behaviour. In order to determine whether close collaboration or the sharing of third party co-authors influences an individual’s overall citing identity, an asymmetrical matrix was developed that included all the co-authors appearing with any of the authors in the consciousness set. These were directly taken from the same articles used to derive citing identities as outlined in Section 5.3.1.

5.3.2.2.1. CO-AUTHOR MATRICES DEVELOPMENT AND VISUALISATION

⁴⁵ ALSCAL is not appropriate for using raw co-citation in this way because does not give an option for treating data as similarities.

⁴⁶ A detailed discussion of cluster analysis techniques was provided in Section 4.7

⁴⁷ The author co-citation cluster analysis solution is given in Appendix E.

The resulting 93 (authors) x 5432 (co-authors) asymmetrical frequency matrix aligns all the Consciousness authors with a combined pool of all the co-authors that have co-authored with all authors across the set. The four authors (NATSOULAS, ROSENTHAL, HODGSON and MARSHALL) that have no co-authorships were not included in the matrix. This matrix allows examination of co-authorships within the identified Consciousness authors set, as well as relationships between Consciousness authors that may not have co-authored directly with another author in the set but with a third shared co-author. From this matrix can be determined the extent to which authors publish with each other as well as the less direct co-authorships that provide less obvious access to their collaborative activity, and the similarity of that activity with other Consciousness authors. A combination of Multidimensional scaling (MDS), cluster analysis and examination of biographical data was used to develop understanding of the similarity of relationships between co-authors and their shared co-authors.

In order to implement the MDS and cluster analyses the asymmetrical matrix was converted to a symmetrical 93x93 author-author matrix. This was achieved by considering counts of co-authors shared by authors in the matrix and then combined with the direct co-authorships within the set. Development of the asymmetrical matrix (which counts the number of times pairs of authors co-author with the same co-authors), means that co-authors will only be included if another author shares at least one co-authorship with that co-author. Authors that co-author with only one other author will drop out of the symmetrical matrix; as such this matrix is not a full representation of all co-authorships but rather relationships with individual co-authors that co-author with more than one of the authors in the set. The result is a measurement of similarity of co-authorships in terms of those that co-author with more than one individual only. Therefore the resulting symmetrical matrix does not differentiate between shared co-author counts and direct co-author counts. This limitation is addressed by consideration of three submatrices: the core matrix (of authors that co-author directly with members of the Consciousness set), the shared co-author matrix (which provides counts for co-authors that are shared by the authors), and the combined core-shared matrix, which is a combination of the first two matrices. This allows for examination of overall similarities between authors. These three matrices together are used to define the levels of representational collaborative activity in the following analyses. Seventeen⁴⁸ authors that did not co-author with any other author in the Consciousness set nor shared any co-authors with them dropped out of

⁴⁸ See appendix D.1 for list

the symmetrical matrix, resulting in a final combined core-shared author matrix of 76x76 authors.

The MDS technique used to visualise the core-shared matrix was the same as that used for the author co-citation matrix detailed in Section 5.3.2.1.1.

5.3.2.2.2. CLUSTER ANALYSIS OF THE CORE-SHARED CO-AUTHOR MATRIX

A comparison of cluster solutions was completed similar to that detailed in Section 5.3.2.1.2. In this case an Average linkage (within groups) cosine 6 cluster solution was identified as describing the underlying matrix the most appropriately from the available options. Appendix F contains the resulting dendrogram.

The results of these analyses were then compared to citing identities in order to explore possible implications of collaborative activity on citation practices across authors' publication careers as captured by the underlying document collection.^{***49}

5.3.2.3. COMMUNICATION CHANNELS AND PARTICIPANTS' REPRESENTATIONAL SPACE - JOURNAL CO-PUBLICATION ANALYSES

Similarity of authors' publication in journals as represented in the records used for the citing identity type analysis are analysed, to determine whether authors that publish in the same set of journals are identified through citing identity types. An alternate though related question is whether communication channels influence citing identity patterns of individual authors. The significance of the question of communication channels for an individual author is raised by Whitley (1980) as they provide a major institutionalising force within a scientific field (see Section 3.4.1). To address these question an initial asymmetric author-journal matrix (97x 1014) was developed from publication records. The development of this matrix involved standardising journal titles using publishers' websites and Ulrich's international periodicals directory for journals. A similarity author-author matrix (97x97) was calculated by developing spreadsheets that counted the instances where authors had published in the same journals. This similarity matrix was then visualised using a Euclidean squared MDS solution that represented the underlying matrix in preference to other solutions, as discussed and demonstrated in Section 5.4.3. A cluster analysis was also completed on the similarity matrix resulting in a Ward's Method cluster solution being chosen for

⁴⁹ This is the same document collection that forms the basis of the citing identities developed in Chapter 4. The timeframe covered is 1990:2002, and consists of documents selected as being written by the author understudy with reasonable assurance. The authors themselves were selected using methods outlined in Section 4.2.5.

reasons of providing a representational identification of the matrix without non-differentiated groups or chaining.⁵⁰

The cluster and MDS solutions were further analysed using a content analysis of the journal titles from the asymmetric matrix. This process involved parsing the titles of the journals and identifying the substantive words or phrases. For titles that were not descriptive of content, for example the journal “BYTE”, the details available from the publishers’ summary were used to identify keywords. The resulting list was only further reduced through identification of the ‘stem’ word (for example “behaviour” is the stem of “behavioural”). This analysis verified the MDS and cluster solution from the matrix and was further developed into a detailed typology of ‘unique journals’.

5.3.2.4. SELF-PLACEMENT OF PARTICIPANTS IN REPRESENTATIONAL SPACE - CITEE MATRIX ANALYSES

Self-placement of individual authors within the boundary specialty of Consciousness was analysed in order to determine whether similarity of authors citing content (the extent of use of similar author’s works) was reflected in citing identity types. The question of whether content similarities at a finer level of detail than those provided by communication channels could be related to citing identities was asked in order to consider the relationship between citing patterns and similarity in research foci.

The similarity of authors’ use of citees work was analysed on the basis of co-occurrence counts between the 97 Consciousness participants. An initial asymmetrical matrix was compiled from the data used to form authors’ citing identities. This resulted in a 97 (authors) x 43232 (citees) asymmetrical matrix. This matrix was then used to calculate co-occurrences of citees for the authors. A purpose written program for generating the symmetrical matrix was used.⁵¹ For example, if AHERN and BAARS have 25 citees in common across the whole citee set, the figure 25 is placed in the symmetrical matrix at the intersection of AHERN and BAARS. This matrix is not a frequency of co-citees but a count of number of similar citees cited by both authors, i.e. the number of citees that these two authors have in common. This calculated resulted in the expected 97 x 97 symmetrical author-author matrix. The MDS PROXSCAL solution, following Leysdesdorff, treated the data as existing proximities (similarities) and ordinal (ties were untied).

⁵⁰ The dendrogram from the analysis is contained in Appendix G.

⁵¹ The code for this program is in Appendix B.2. This program was co-developed with Andrew Beehag

A cluster was also completed. The possible solutions were compared according to the methods outlined in Section 4.2. The Average linkage (within groups) 10 cluster cosine solution was the best available according to these criteria; however this solution is not optimal as this solution includes 3 isolates.⁵² The MDS visualisation, the cluster analysis and biographical information of the authors were used in combination to attempt to overcome the methodological limitations for the analysis of citing identities.

5.3.3. CITING IDENTITY TYPE ANALYSES

The means of generating citing identity types using cluster analysis has been presented in Chapter 4.

The citing behaviour of individuals was analysed through measurement of the individual features of their citing profile. It was expected that aspects of an individual authors citing distribution would potentially have a relationship to the placement of that individual in terms of their citing identity type. This also examines the question as to whether one aspect of the distribution is dominating the formation of citing identity types over any of the others. The aspects analysed are productivity (simple and fractional), unicitations (unicites), recitations (recites), self-citing and length of publication career as represented in the ISI citation indexes.

The data used for these analyses was extracted from the publication records collected for the citing identity type analysis, see Section 5.3.1. Biographical information was also used for age of individual and to develop a keyword based typology of research field affiliation as much as possible as given by the researcher themselves, see Section 4.2.4.

5.4. RESULTS - RESEARCH PERSONAS IN REPRESENTATIONAL SPACE AND CITING IDENTITY TYPES

5.4.1. INTELLECTUAL PLACEMENT OF PARTICIPANTS BY OTHERS - AUTHOR CO-CITATION ANALYSIS (ACA)

Following Leydesdorff (2005 preprint) a PROXSCAL MDS 2x2 dimension solution was developed, where the raw co-citation matrix was treated as a similarity ordinal matrix, with ties untied (see Section 5.3.2.1). This procedure resulted in a solution that had low normalised raw stress at 0.05, a high Dispersion accounted for at 0.95, and Tucker's coefficient of congruence at 0.98, indicating the appropriateness of this MDS solution

⁵² The dendrogram for this solution is provided in Appendix H.

in terms of good fit to the underlying data matrix. The visualisation of the PROXSCAL solution with co-citation clusters overlaid is presented in Figure 5.1.

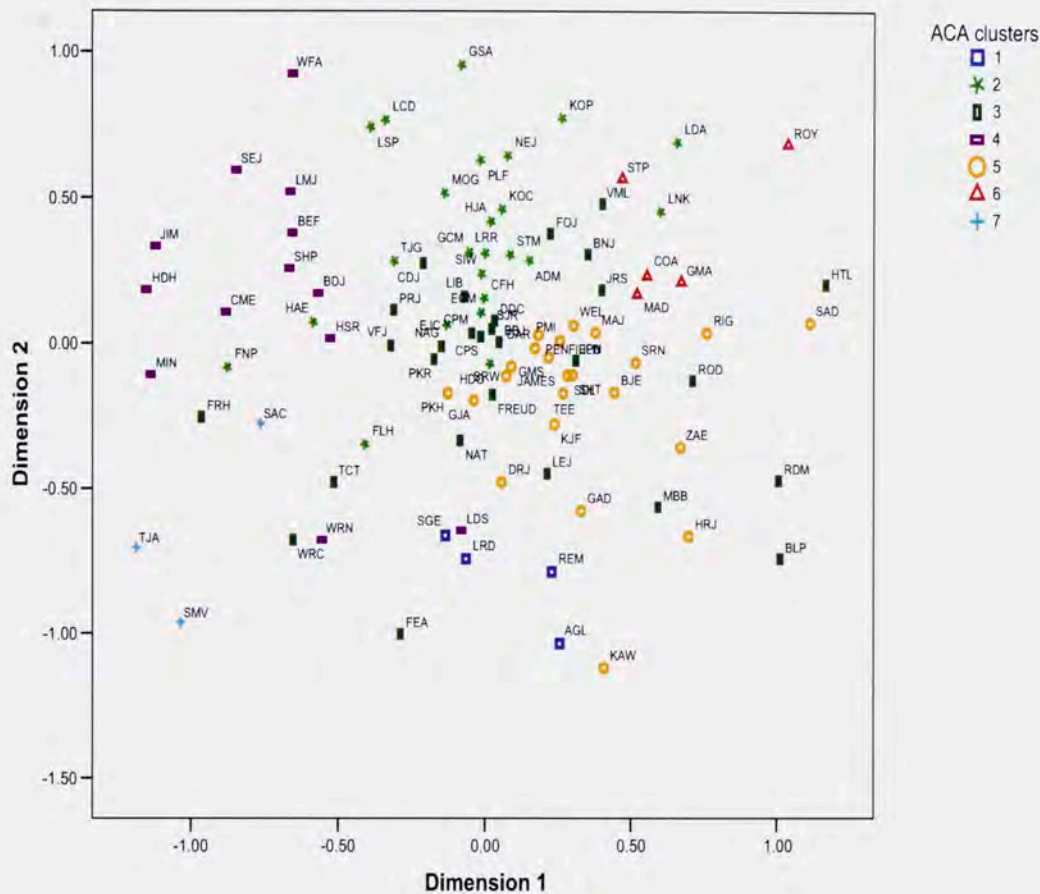


Figure 5.1 MDS PROXSCAL solution for Author co-citation matrix with author co-citation clusters overlaid

Further evidence in support of this particular MDS PROXSCAL solution is found when the dimensions are compared with co-citation counts and totals from the underlying co-citation matrix. Authors that have higher co-citation counts and totals are placed in the central positions in the MDS solution, whereas authors with lower co-citation totals and counts are placed in the periphery of the solution. This is demonstrated in Figure 5.2 where it can be seen that authors with that have the highest co-citation counts only appear in the central position on Dimension 1(0.00), whereas authors that have lower co-citation counts fan out into the peripheral positions along the dimension. This pattern holds from dimension 2 though not quite as consistently for co-citation counts and totals as it does along dimension 1.

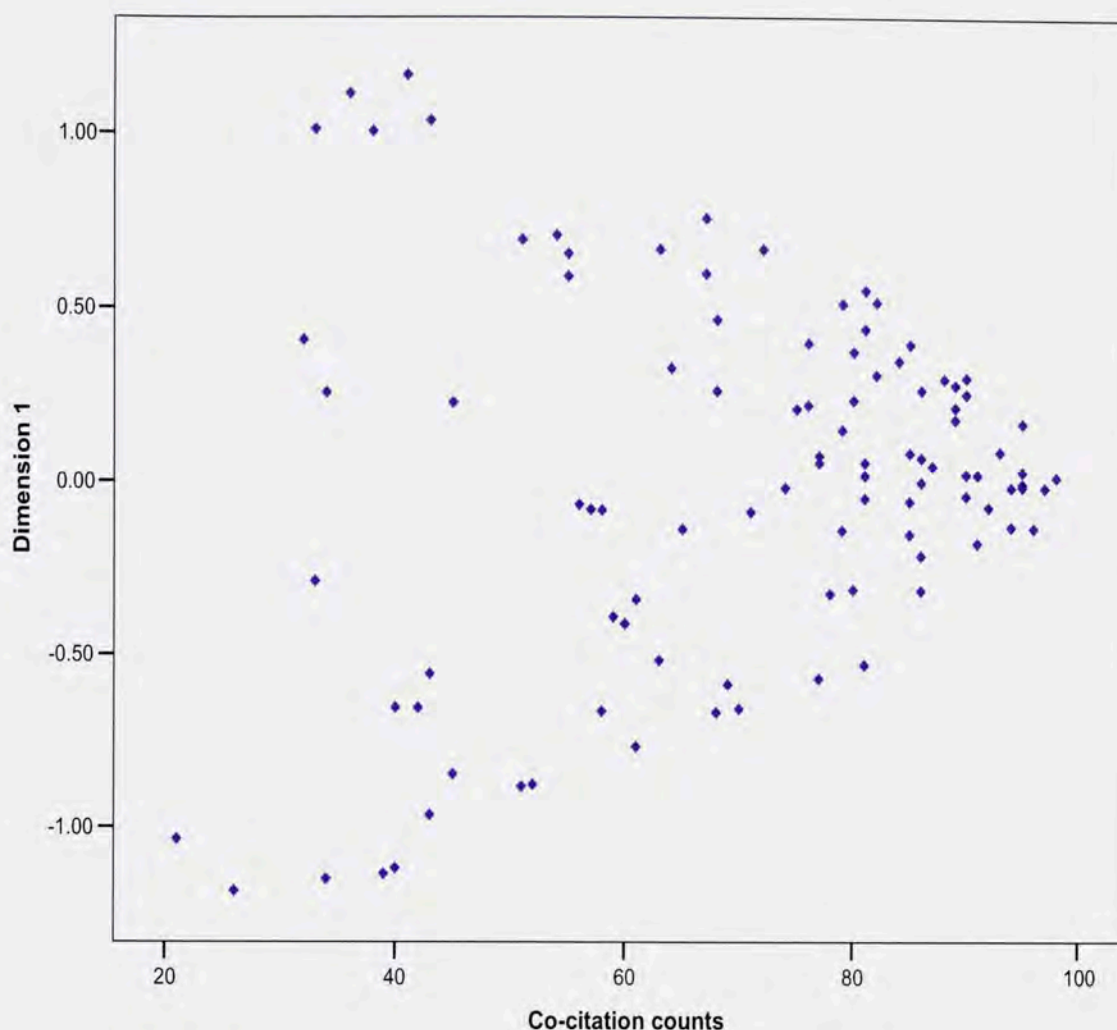


Figure 5.2 MDS PROXSCAL solution dimension 1 in relation to co-citation counts from co-citation matrix

Aligned with co-citation totals and counts is the relationship between the authors age and their placement in the field. Figure 5.3 demonstrates the interesting tendency for the most mature researchers (founding fathers as it were) in the field to also be almost without exception the most central. Dimension 2 demonstrates this most strongly while dimension 1 also demonstrates this tendency but not so definitively. GIUSEPPE MORUZZI (MOG) is the exception; he is an Italian neurophysiologist who examined “the electrophysiological relationships between neuronal systems in the brain and epilepsy”. His work is not centrally co-cited to the same extent to other authors with similar extensive histories. However others in this group are most definitely central to the development of the scientific study of Consciousness, which this MDS solution captures and is supported through examination of sources describing the development of the field. For example FRANCIS CRICK (CFH) is found in this group. CRICK’s book “The Astonishing Hypothesis” (Crick, 1993) establishes the groundwork for the scientific study of Consciousness by arguing for a programme of study to examine the Neural correlates of Visual perception. CRICK is most famous for his work on DNA with James

Watson, for which he won a Nobel Prize in 1962, after which he turned his attention to supporting the establishment of the scientific study of Consciousness. A long range and popularistic account is captured in the following quotation.

A major shift in the neurosciences occurred in the 1990s: the topic of consciousness and its relation to brain function has become a respectable topic that many neuroscientists take seriously. Prior to the 1990s, few neuroscientists spoke of consciousness, and even fewer would be bold enough to try to approach the topic scientifically. Consciousness was not considered to be a topic that was amenable to the methods of science. The tide change in the neuroscientific community of the 1990s is largely due to outspoken scientists such as Nobel-laureate Francis Crick, and philosophers such as David Chalmers. While neuroscience has not yet solved the mind-brain problem in terms of coming up with an NCC, to many in the field, the next decade looks promising. (*Mind-body problem*, 2005)

SIGMUND FREUD and WILLIAM JAMES are also found in the group identified in Figure 5.3. The unifying feature of the significant majority of the authors identified as belonging to the group labelled in Figure 5.3 is that their work forms the foundational basis of the scientific study of consciousness. Another example is JOHN ECCLES (EJC), the British neurophysiologist, who speculated in 1986 “that synapses in the cortex respond in a probabilistic manner to neural excitation, a probability that could well be governed by quantum uncertainty given the extremely small size of the synapsis' "microsite" that emits the neurotransmitter. If this is true, ECCLES speculates that an immaterial mind (in the form of "psychons") controls the quantum "jumps" and turns them into voluntary excitations of the neurons that account for body motion” (Scaruffi, no date). This work is a foundational element in the current debates of the role of quantum processes in neural activity that forms the basis of consciousness. ECCLES was also a prominent advocate of the dualist position in the field, and as such forms a focal point for both those who agree with his stance in some sense and also for those who present arguments in opposition (Dennett, 1993).

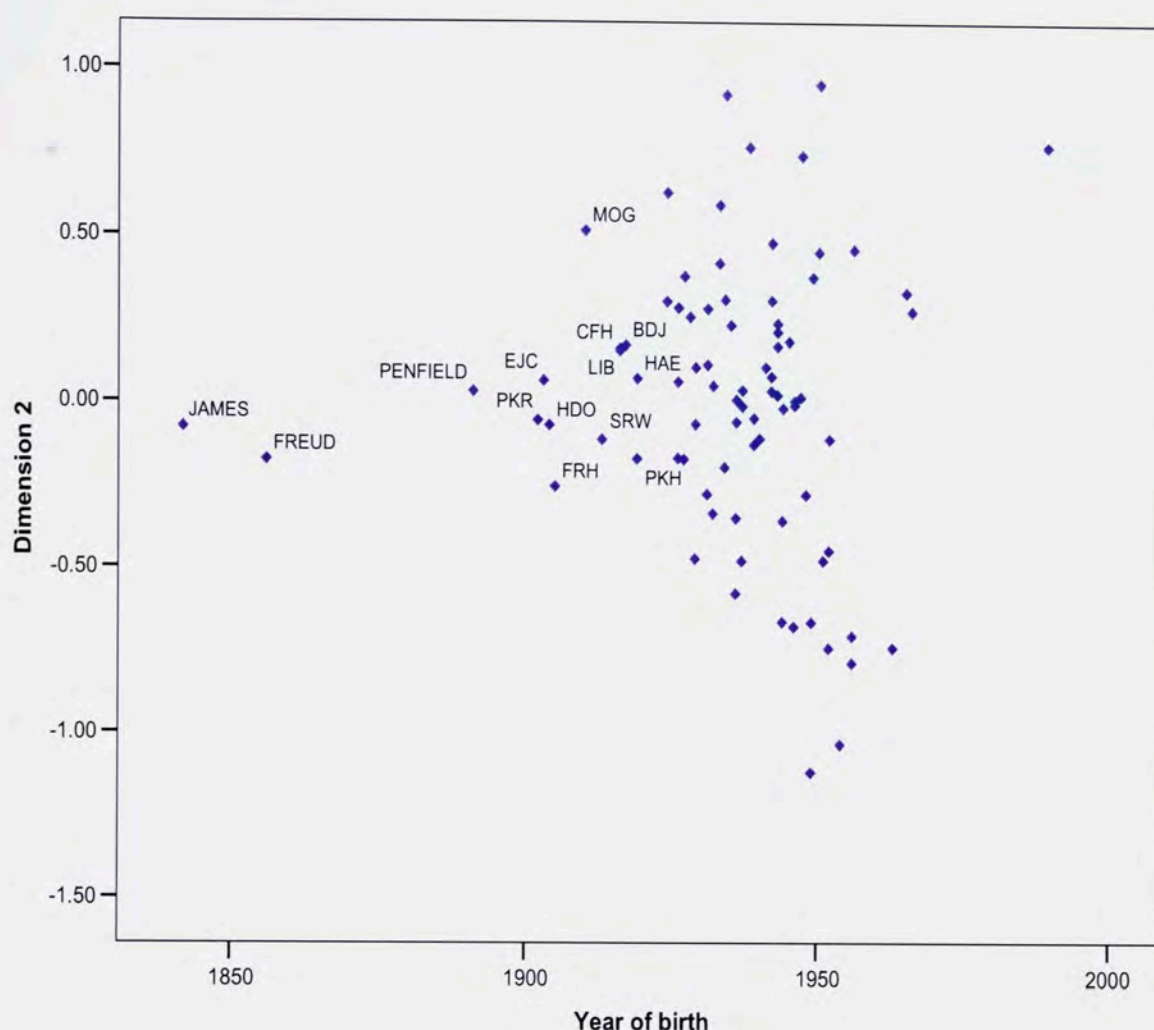


Figure 5.3 MDS PROXSCAL solution dimension 2 and authors year of birth

While it is the case that authors with the longest surviving co-cited work are placed, with only one exception, in the central position in the MDS solution it is also the case that it is not only these authors that are centrally placed (Figure 5.3).

5.4.1.1. INTELLECTUAL PLACEMENT OF PARTICIPANTS BY OTHERS AND CITING IDENTITY TYPE

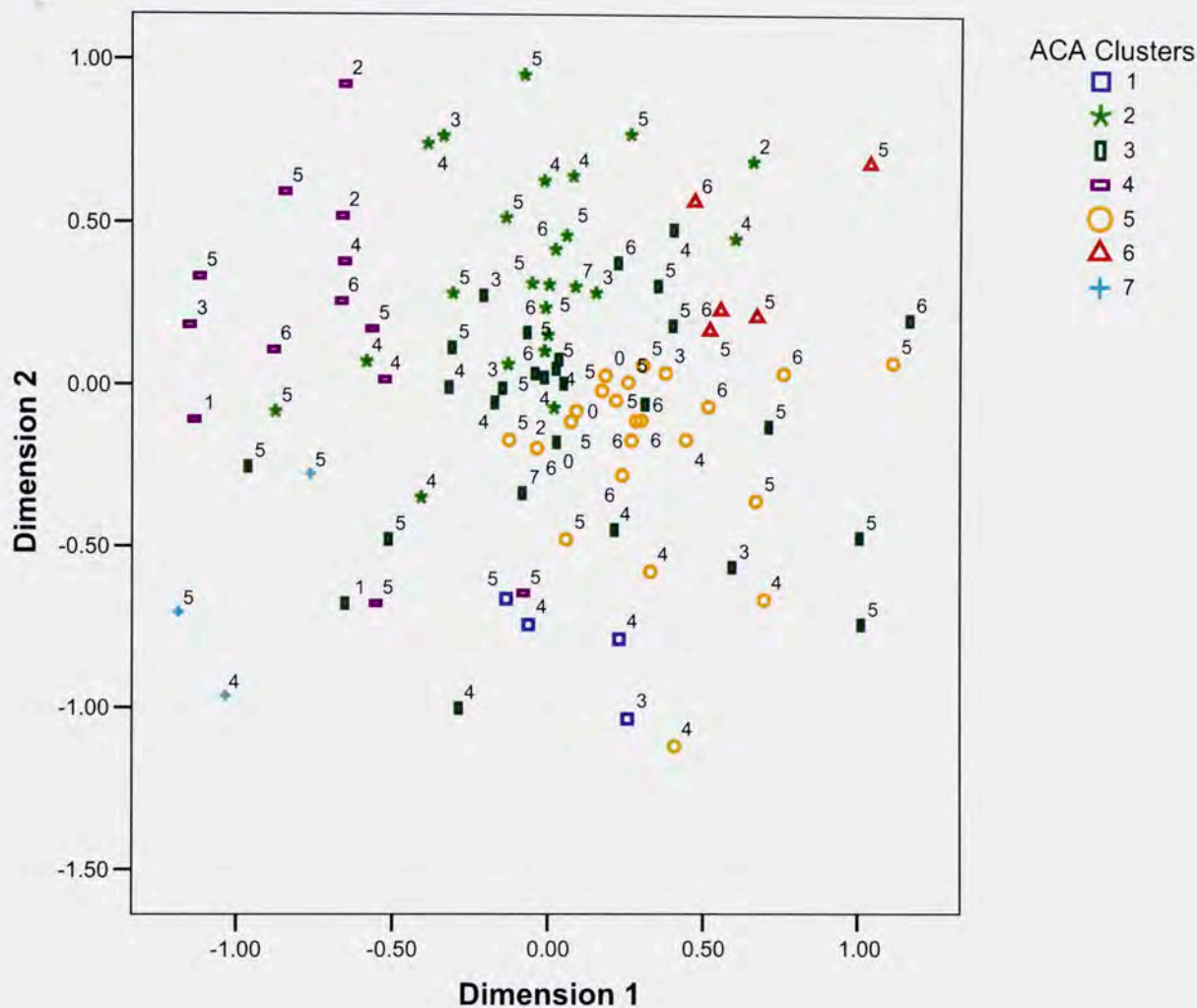


Figure 5.4 MDS PROXSCAL solution for Author co-citation matrix with author co-citation clusters overlaid with citing identity types

When the citing types are overlaid on the Author co-citation analysis no clear pattern emerges (Figure 5.4). All of the citing identity types are distributed across the map.⁵³ However the right hand side of the map (x-axis >1) can be seen to be dominated by citing identity type 5, though this is not without exception. Citing identity type 6 can only be seen to be present on the left hand side of the map (x-axis <1), and similarly citing type 4 is also much more frequent within this region. There is a notable exception in citing identity type 4, MICHAEL SATARIC (SMV) appears at the far left hand side of the map very near JACK TUSZYNSKI (TJA) (citing type 5) who is his relatively frequent co-author (13 times within ISI data set). RICHARD WATT (WRC) (citing type 1) and ROGER WALSH (WRN) (citing type 5) also appear in very close proximity to one another on this

⁵³ The “o’s” represent authors that have been removed from the analysis of citing identity types due to insufficient representation of career publications within ISI time frames (see Section 4.2.5).

map, these authors however do not co-author together. There is a large distance between the two authors from citing type 7, THOMAS NATSOULAS (NAT) and MIRCEA STERIADE (STM).

As can be seen in the distribution of citing identity types across the boundary specialty of Consciousness intellectual placement has no clear relationship to citing identity types of authors. Even where the MDS solution represents central placement within the specialty citing identity type patterns are not evident.

5.4.2. COLLABORATIVE SIMILARITY BETWEEN AUTHORS BASED ON CO-AUTHOR AND SHARED CO-AUTHOR MATRIX ANALYSIS

Figure 5.5 is a MDS visualisation of the combined core-shared co-author symmetrical similarity matrix. This solution was developed using the PROXSCAL program in SPSS. Combined core-shared co-authorships are treated as an existing proximity matrix, that measure similarities, these similarities are ordinal and untied; a 2 dimensional solution is used as further detailed in Section 5.3.2.2. Normalised raw stress for this solution is appropriately low at 0.007, indicating that the misfit of solution to the data is low while dispersion accounted for is high at 0.99 and Tucker's coefficient of congruence measure the fit is also high at 0.99. Further supporting applicability of this solution, in that it provides a good representation (with limitations discussed below) of the underlying similarity matrix, is in accord with recommendations for use of this MDS methodology by Leydesdorff and Vaughan (2005 preprint).

In order to further develop an understanding of collaborative similarity amongst the authors a cluster analysis was performed again using SPSS. The solution presented is an Average linkage (within groups) cosine solution with 8 clusters. This solution provided differentiation between clusters but did not produce the one overly large cluster that was formed (containing approximately 50% of authors) using a more direct Squared Euclidean solution; this was combined with an undesirable chaining effect that also was a feature of that solution. The 8 cluster Average linkage (within groups) cluster solution is overlaid on the MDS solution in Figure 5.5 for comparative assessment.

cluster 6; however in the MDS solution he is found in a more central position (0, 0.9); this position can be seen to be somewhat removed from the majority of the other authors placed in cluster 6 on the MDS solution. When the underlying combined core and co-author similarity matrix is examined the only link FLANAGAN has to any authors or shared co-authors in the combined set is via a shared third party co-author with DONALD HEBB (HDO), to whom he is more closely placed in the MDS. In this instance the level 8 cluster solution does not provide adequate discrimination to capture FLANAGAN's co-authorship activity. This seems to be relatively common when the linkages to the overall set in terms of co-authorships are low.

An instance where neither MDS nor cluster seem to capture relationships is EUAN SQUIRES (SEJ). SQUIRES was Head of Mathematical Sciences Department, University of Durham, England, before he died in 1996. One of his research areas was squarely within the area of quantum theory and consciousness; however when the underlying matrix is considered there is no reason why SQUIRES should be placed in co-author cluster 6. SQUIRES has one shared co-author with BOHM (BDJ, co-author cluster 4) but is not placed within the same cluster as BOHM nor does he appear in close proximity to him in the MDS (SQUIRES at co-ordinates -0.90, -0.26 while BOHM 0.13, 0.61).⁵⁴

However there are also many instances where the co-author cluster analysis and the MDS solution coincide and capture the significant relationships between authors within the underlying combined core author-shared co-author matrices. Groups that work together more closely, or author pairs that work together more closely and more frequently as apparent in the underlying matrices, appear to be closely placed in the MDS solution. For example, within the marked box in Figure 5.5 is a group of authors that also form a group in the core author network, in other words they all directly co-author with one another. This group is centred around the co-authorship activity of STUART HAMEROFF who is Director of the Centre for Consciousness Studies at the University of Arizona (the group primarily responsible for organising the Annual Conference series "Scientific studies of Consciousness"). A close alignment of cluster 6 in the MDS solution and the same group of direct co-authorships is apparent when core author networks (Figure 5.6) are compared with the combined MDS and cluster solution (Figure 5.5).

⁵⁴ rotating solution may help with these instances however others may concurrently be comprised

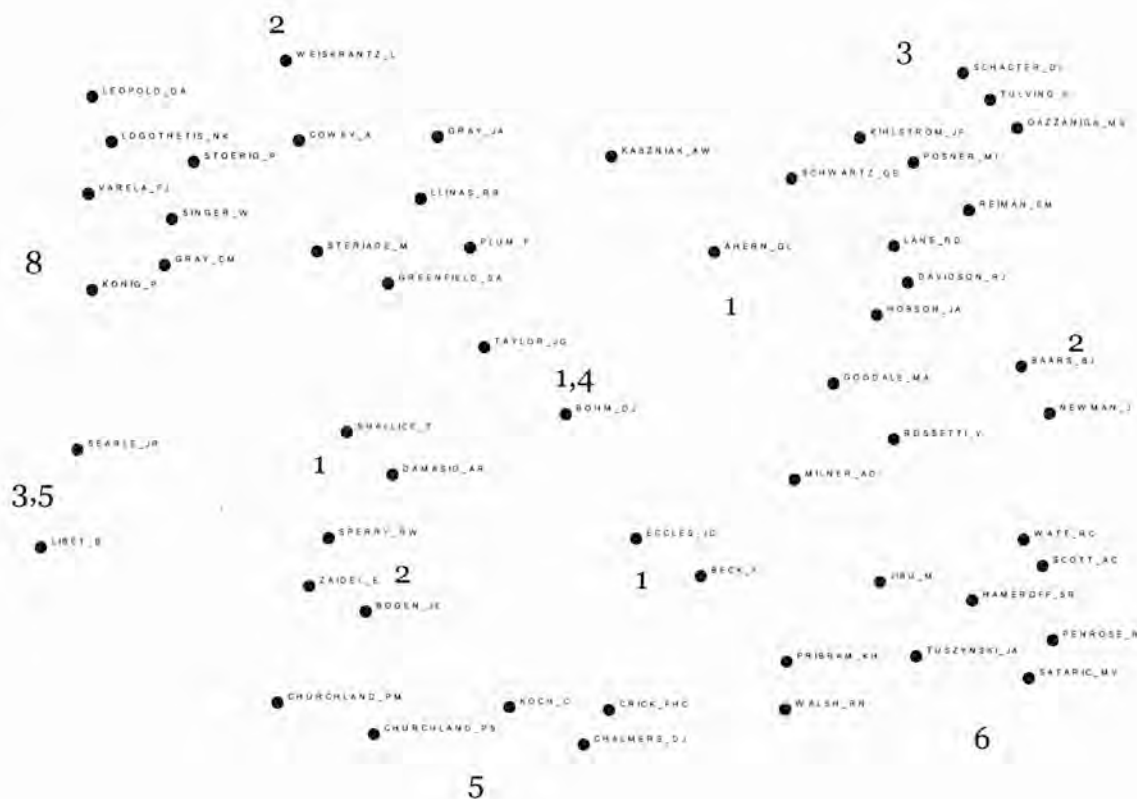


Figure 5.6 Core author direct co-authorships networks. Numbers indicate the corresponding assignment of authors to co-author clusters.⁵⁵

Examination and comparison of co-author clusters, the underlying direct core co-author, shared co-authors and combined matrices combined with data from their biographical data indicate that in this instance the combined MDS solution and cluster analysis representation do capture the close collaborative relationships amongst members of co-author cluster 6 as a detailed example. Five of the twelve authors in cluster 6 work focuses on the connections of quantum theory and consciousness, including HAMEROFF, JIBU, TUSZYNSKI, SQUIRES and PENROSE. Three more authors of in cluster 6 focus their work in the related area of nonlinear systems modelling of biophysical processes including consciousness (FROHLICH, SCOTT and SATARIC). While RICHARD WATT works directly with HAMEROFF at the University of Arizona, his theoretical work is closely aligned with HAMEROFF and the 10 papers he writes that do not list HAMEROFF as a co-author are focused on patient monitoring and experimental work with Anesthetics. KARL PRIBRAM's work is focused on cerebral function as it relates to psychological processes. He co-authors directly with HAMEROFF, JIBU and

⁵⁵ These core co-author networks were developed from a 180x180 symmetrical author-author frequency matrix that includes only those co-authorship relationships between authors within the Consciousness author set. The visualisation was produced using the NetDraw program that is part of the UCINET software package. The links between author represent direct co-authorship relationships, thickness of the line indicating the strength of this relationship.

WALSH, though also shares co-authors with 16 others authors including BOHM, DAVIDSON, FROHLICH, and KILHSTROM, which is indicated by his position in the MDS: while still close to cluster 6 he is moving toward the centre and other authors. The two authors whose membership in this cluster is somewhat inexplicable are LABERGE (LSP) and FLANAGAN (FOJ). FLANAGAN only has one shared co-author with HEBB, who is not in cluster 6 but cluster 1; in the MDS solution he is not placed with the rest of cluster 6 but rather closer to HEBB. LABERGE has one shared co-author with HEBB (cluster 1) and LIBET (cluster 5); however like FLANAGAN in the MDS solution he is not placed with the rest of cluster 6 but rather closer to HEBB and closer still to LIBET.

Representational collaborations as they reflected in the publication set of Consciousness participants appear to be formed by a diverse range of researchers that can be seen to come together within specific problem foci. Each of the participating authors brings a perspective to the problem that is sometimes highly integrated with others members of the collaborative group but in the majority of cases the approach contributed is differentiated from other members by a differing methodological approach or sub-problem focus. One of the larger co-author clusters (1) is detailed to provide illustration of this collaborative activity.⁵⁶

Co-author cluster 1 includes seventeen individual authors from the Consciousness author set. FRIEDRICH BECK (BEF) demonstrates the most content differentiation from his fellow members in this cluster. BECK's early work and some of his later work is strongly focused on purely physicist problems relating to heavy ion-collisions, electron scattering and quantum hadrodynamics, and relativist models of process. In the early 1990's however, BECK's attention partially turned to the mind-body problem, first publishing a paper in 1991 on representation in the relativistic mind-body problem, and then going on to co-author a paper with JOHN ECCLES (EJC) published in 1992 entitled "Quantum aspects of brain activity and the role of Consciousness". From this one co-authorship ECCLES forms BECK's link to this cluster. Examination of the asymmetrical matrix revealed that BECK also has 1 co-authorship tie with MILLER GA, and MILLER GA has one tie each with HOBSON & DAVIDSON, but from further examination of the underlying papers concerned it seems very unlikely that the shared MILLER GA is indeed one and the same person. Apart from separate affiliation information one MILLER appears to be a psychologist and the other is a physicist; an instance where shared co-

⁵⁶ Details are derived for the most part from listings of research interests provided by the researchers themselves made publicly available via the internet. In instances where lists provided by the author themselves are unavailable a combination of what biographical data is available with frequency of keywords used in the articles in the set have been used. See Section 4.2.4.

authorship analysis breaks down even with concerted attention to isolating the appropriate individuals. For these reasons BECK's relationship to the cluster cannot be considered strong, which can be identified by his peripheral placement on the MDS in relation to other members of the group.

ECCLES however has stronger ties sharing co-authors with ALFRED KASZNIAK (KAW) and DANIEL LEVINE (LDS). The work that is captured in this data set is only ECCLES' very late contributions, in fact only after he had formally retired in 1975 (he was born in 1905 and died in 1997). The main focus of this work is however in keeping with the general problem area of higher-order functions and their relationship to psychophysiological characterisations that forms a consistent focus for this whole cluster. ECCLES' work however is much broader in scope in comparison to other members of this cluster. Primarily ECCLES is something of a precursor to the work of the other members of this group arguing directly for neurophysiological examination of behaviour, but his studies also encompass specific aspect in relations of higher-order functions such as learning, speech and the creation of self. HEBB (HDO), another elder statesman, also provides a more foundational perspective than the work of the other members of this cluster, presenting a general theory of behaviour that attempts to bridge the gap between neurophysiology and psychology. His specific sub-problem areas within the Higher-order function problem area are visual perception (particularly in his early work), emotion and intelligence.

The remaining relationships between authors in cluster 1 demonstrate a closer homogeneity, both in terms of their direct co-authorships and their shared co-authors. These authors without exception from one approach or another focus on the problem area of higher-order functioning, specialising in particular aspects and examining these aspects through various methodological approaches. Nine of the remaining fifteen authors focus on emotion (not including HEBB who also includes this research interest), all applying psychophysiological methods such as functional imaging, neurobehavioural clinical studies or neural dynamics modelling. Studies associating function variation with neurophysiological disorders are also featured in this cluster. The remaining higher-order functions include visual perception, language, learning, memory, anxiety, sleep and intelligence.

The direct co-author relationships in cluster 1 represent a combination of approaches to the studies of these higher-order functions. For example ALFRED KASZNIAK (KAW) and GARY SCHWARTZ (SGE) co-author 7 papers together and share 16 third party co-authors. KASZNIAK and SCHWARTZ have reasonably commensurate career lengths

receiving their doctorates in the years 1976 and 1971 respectively. KASZNIAK's research interests are listed as

Clinical Neuropsychology of Aging, Dementia, Self-Awareness, and Emotion.... Specifically, my laboratory and clinic research currently involves four different, although related, domains of interest: (1) Neuropsychological aspects of aging; (2) Neuropsychological aspects of age-related disorders of the central nervous system; (3) The neuropsychology of consciousness and self-awareness; and (4) Brain systems in emotion. (Kaszniak, No date)

demonstrating a central focus of aging and associated functionality, with particular interests in clinical approaches to studying higher-order functions including emotion and memory. SCHWARTZ also has a clinical interest in psychophysiological process in relationship with ailments, also studying higher-order functions including emotion and hypertension. The work they co-author for the most part draws together the study of emotion as it relates to memory and aging. SCHWARTZ does not publish on age related ailments in the absence of KASZNIAK but rather on environmental impacts on memory and other higher-order functions in various age related populations. Each member of the co-authorship team combines their area of interest in this team in a way not seen in papers either authors produces with other co-authors.

Diversity of composition of authors within individual co-author cluster and groups extends into the extent of experience represented within the groups. Figure 5.7 indicates this diversity of composition by demonstrating the association of authors with longer research careers with those that have shorter career lengths to date.⁵⁷

⁵⁷ Year of highest degree, in most cases the PhD, has been used as an estimate measure of length of career, with a majority of authors having a commensurate length of publication careers. In cases that authors received their highest degree significantly prior to 1972, 1974 or 1980 (the beginning dates for the ISI Citation index data available – 1972 for Social SciSearch the Social Science Citation index, 1974 for SciSearch the science citation index, and 1980 for Arts & Humanities Search the Arts & Humanities citation index) their publication record begins within this study at this time (ie the maximum timespan subject to access restrictions). For younger researchers receiving their highest degree relatively more recently their first publication is within two-three years before or after the year their highest degree is granted.



Figure 5.7 Direct core co-author networks labelled with year of receipt of highest degree

As can be seen in Figure 5.7 each individual co-author network contains members that vary considerably in terms of length of their research careers as indicated by the year in which they received their highest degree. Members with longer careers are almost invariably connected with less experienced colleagues.

5.4.2.1. COLLABORATIVE REPRESENTATIONAL SPACE AND CITING IDENTITY TYPE

5.4.2.1.1. CITING IDENTITY

Given the diversity in co-authorships and shared co-authors represented in the Consciousness boundary specialty perhaps it is not surprising that no relationships were found between similarity of co-author and shared co-authors and citing identities defined either through citing type clusters or the Gini measure itself. As can be seen in Figure 5.8 citing types are dispersed throughout the MDS solution with no particular patterns emerging.

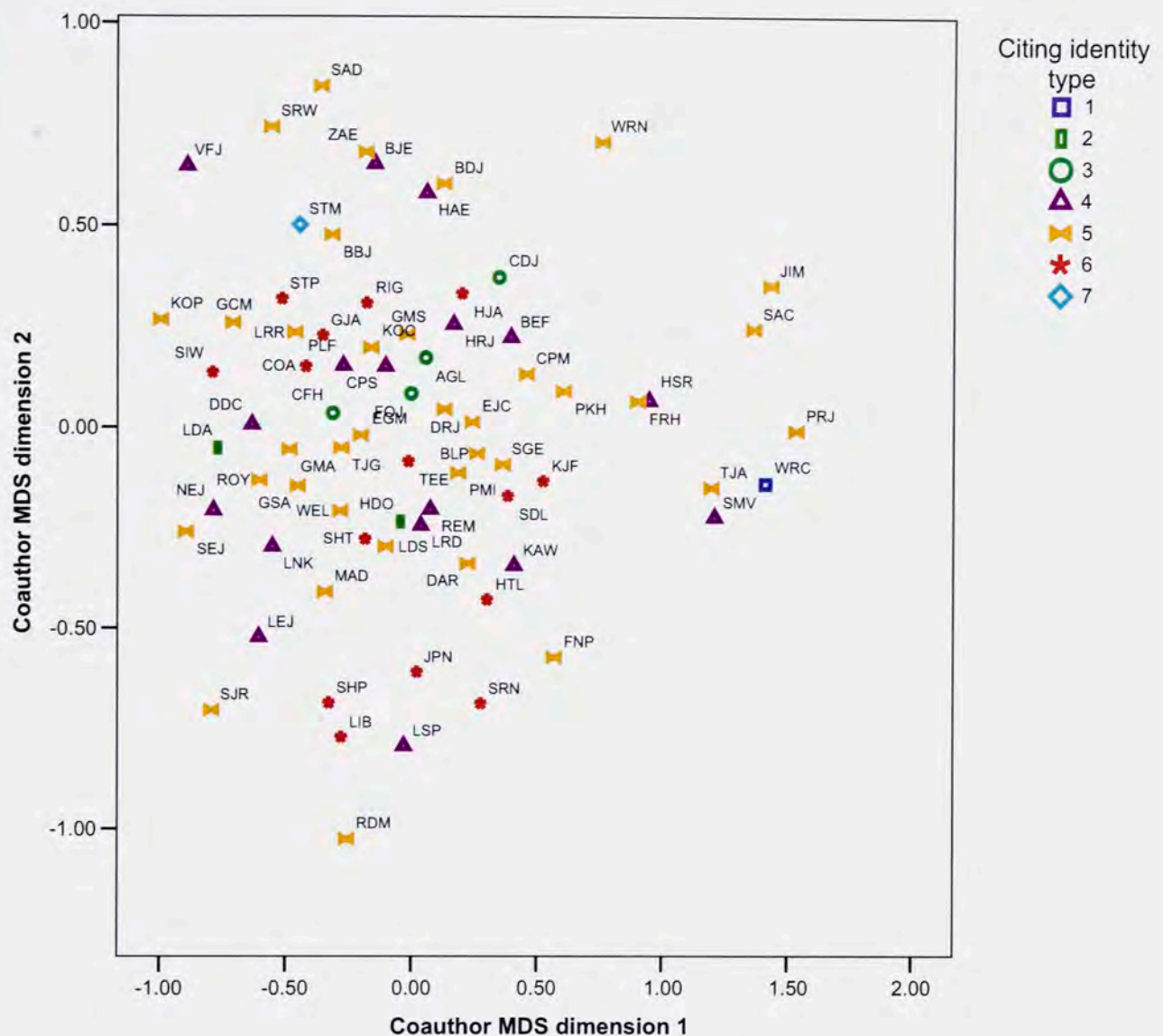


Figure 5.8 MDS PROXSCAL solution overlaid with citing identity types

Another feature evident by considering the citing identity types overlaid on the MDS co-author solution is that there is a clear predominance of citing identity type 5. Citing identity type 5 is a very large proportion of the author citing types, having a membership of 41 of the 97 Consciousness authors, so it is not surprising that the type's dominance is apparent here. However, some of the less populated citing types have proportionately been more drastically reduced due to the combined fact that some groups are smaller and have authors that co-author less frequently. As an illustration citing cluster 6 is populated by individuals that all co-author with other members of the Consciousness field directly or they share co-authors. Citing identity types 1, 2 and 7, though smaller groups to begin with, have half their membership disappear once co-authorships are considered. Citing identity type 3 has 40% of its authors that do not share co-authors or participate in co-authorships with other Consciousness researchers. This is where some of the limitations of a multidisciplinary case field

become apparent. Citing identity type 3 is predominantly comprised of Philosophers, who in general practice do not co-author as much as their colleagues with other disciplinary affiliations. Therefore especially for researchers such as Philosophers, collaborative activity needs to be considered using alternative methods than assessing co-authorships alone; appropriate methods could include interview or survey. However their utility in this regard comes down to asking authors, or having access in some way to their direct practices that do not involve relying solely on their publication records, especially given the other important limitation of journal article publications only.

In light of these considerations Figure 5.9 includes co-author cluster 0, this category does not appear in the MDS solution as the authors included within this cluster either do not co-author at all (4 authors) or they do not share co-authors with others from the Consciousness author set (17 authors). However in line with the diversity found with citing identities in the other co-author clusters, this cluster includes authors across the whole range of citing identity types.

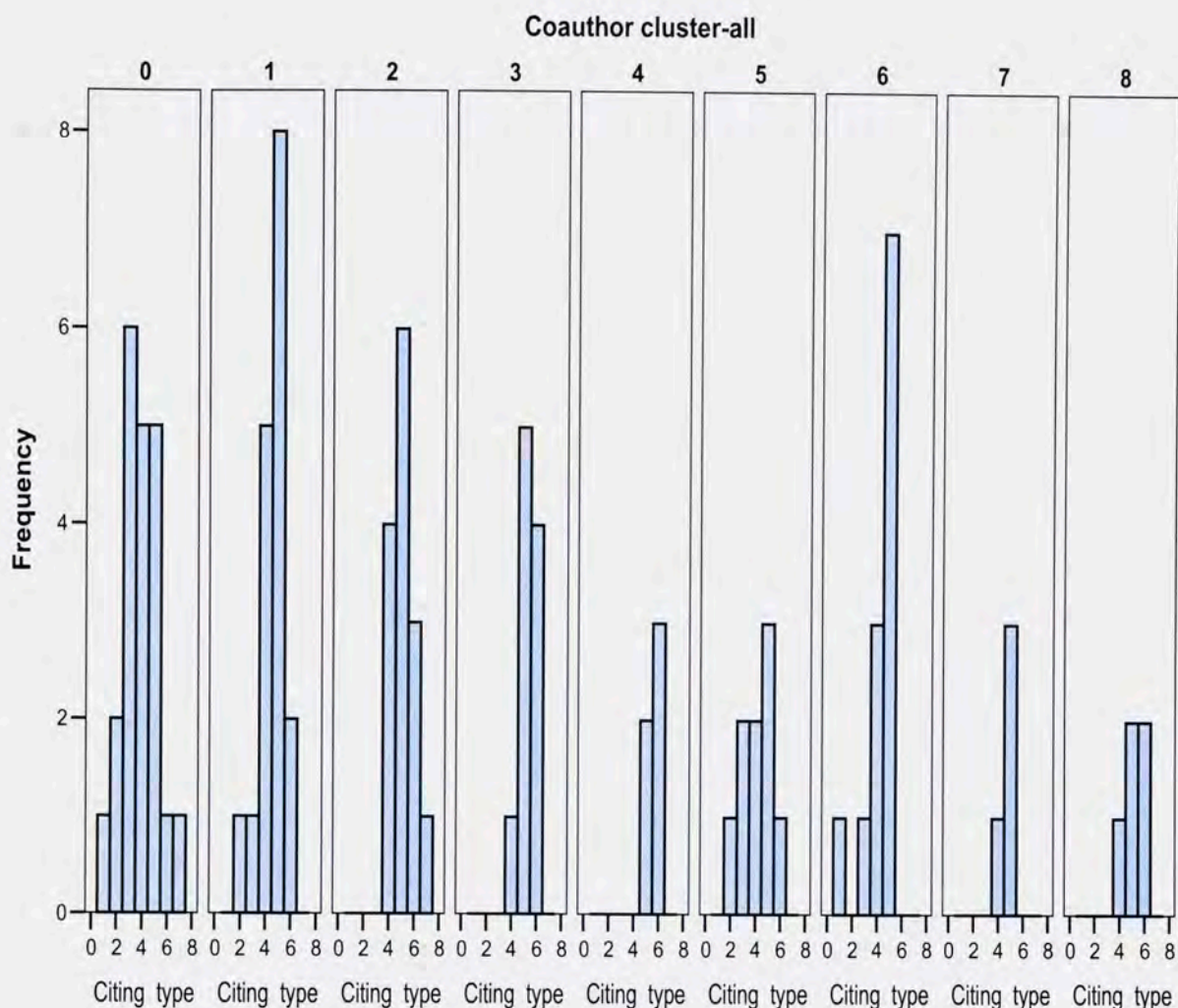


Figure 5.9 Frequency of occurrence in co-author cluster according to citing identity type

Figure 5.9 does demonstrate a weak tendency for lower citing identity types (citing type 4 and below) not to be members of co-author clusters 2, 3, 4, 7, and 8,. When co-author cluster 0 is considered authors with either no co-authors or no shared co-authors, those that are not collaboratively integrated into the Consciousness author set still belong to the whole range of citing identity types. NATSOULAS (NAT) is an interesting case in point. According to his citing identity type (7) NATSOULAS has a very concentrated citing distribution, he also cites comparatively often (4852 total cites) and has a moderately high overall productivity (102), especially when compared to the next most productive author that has no shared or direct coauthors, DAVID ROSENTHAL (ROD) at 31 publications. When NATSOULAS is examined in light of his overlapping attributes he clearly has a very distinctive research career. NATSOULAS is a theoretical psychologist by disciplinary affiliation, his productivity levels conform in general terms to other psychologists; however strong individualism is indicated by his highly concentrated journal publication activity, high self-citation rates and no co-authorships. In a limited

though fundamental sense NATSOULAS' appearance in co-author cluster 0 combined with his citing identity indicate an individualistic publication and scientific career.

Other co-author clusters are populated with a range of citing identity types, no particular pattern of citing identity type is strongly associated with any of the co-author clusters. Co-author cluster 6 includes RICHARD WATT (WRC), an author with citing identity type 1. WATT shares co-authors with his colleagues at the University of Arizona, but only directly co-authors with STUART HAMEROFF (HSR) from this group. STUART HAMEROFF is an Anesthesiologist, who is the director of the Center for Consciousness Studies at Arizona; his work focuses on the role quantum explanations play in understanding Consciousness. As director of a prominent organisation for Consciousness studies it is not surprising to find HAMEROFF at the centre of the network diagram and the connecting individual (Figure 5.6) in a group of closely connected direct co-authors within the field; this for the most part comprises cluster 6. WATT by comparison publishes far less in the field of consciousness; in fact his only directly related work is with HAMEROFF, WATT's other work that is not with HAMEROFF involves experimental work in administration of anaesthesia. WATT is an electrical engineer by training. The analysis has highlighted that while co-authors clearly share publications in the data set being examined, they often are co-authoring with those from diverse backgrounds. From this perspective the dispersion in citing identity types within the clusters and across the MDS is perhaps related to the underlying nature of collaborations and a range of citing types would be appropriate within a co-authorships group.

Similarity of co-authors or integration into fields' collaborative activity does not correlate with similarity in citing identity distribution. It is possible (if not likely) that this is due to the nature of collaboration (or at least as it is expressed in this boundary specialty). Collaborations most frequently occur between those from different backgrounds and experience levels, but with similar focal problems. Generally the pattern includes relatively few central individuals within each cluster connecting and collaborating with many others. However if this is the case, dispersion of citing identity types does reflect to some extent this diversity in contributing authors backgrounds or experience levels. Due to limitations of the data set, this question cannot be adequately addressed in the context of this work; however preliminary investigation of potential patterns arising from diversity within sub-groups could be informative, taking experience levels as a starting point.

5.4.2.1.2. STUDENT/MENTOR CO-AUTHORSHIPS AND CITING IDENTITY

All citing identity types and co-author clusters represent a diverse range of experience levels; this is verified by both the Core network of direct co-authorships above (Figure 5.7) and a similar analysis across the field both according to the MDS solution and the co-author clusters. In order to examine possible patterns in terms of available data, citing identities between student-mentor pairs were examined. Both authors are restricted to being one of the case authors and need to have formed such a relationship within the field as defined by this study. Table 5-1 outlines the known student-mentor relationships within the Consciousness data set, the relationship between the authors is described and authors are identified with both the year of their highest degree and their citing identity type.

Table 5-1 Table of known Student-Mentor relationships within the Consciousness data set

Author	Year	Citing identity type	Mentor	Citing identity type	Year	Relationship	Collaborate in timeframe
GMS	1964	5	SRW	5	1941	PHD supervisor	N
GMA	1969	5	WEL	5	1953	Post-doc supervisor	N
GCM	1986	5	SIW	6	1972	Post-doc supervisor	Y
GSA	1977	5	LRR	5	1965	Post-doc supervisor	Y
KOC	1982	5	CFH	3	1954	Mentor	Y
KOP	1989	5	SIW	6	1972	Post-doc supervisor	Y
LDA	1997	2	LNK	4	1985	Post-doc supervisor	Y
RIG		6	MOG	5	1933	PHD supervisor	N
SDL	1981	6	TEE	6	1957	PHD supervisor	N

As can be seen in this table, three student versus mentors relationships consists of a higher (increasingly concentrated) citing distribution for a mentor than for a student. The three pairs are LEOPOLD (LDA) and LOGOTHETIS (LNK), KONIG (KOP) and SINGER (SIW), and GRAY (GCM) and SINGER (SIW). It is potentially interesting to note that for these three cases the time of highest degree received by both pair members is within, or close to within, the timeframe covered by the dataset. For all the other pairs that have either mentors having lower or equal citing concentration, at least the Mentor’s publication career publication representation is reduced according to the extent their career is captured. The timeframe covered by the available ISI set does not allow for representation of any of their early career publications; in the case of CRICK (CFH) there is a 18 year gap between receipt of highest degree and the earliest possible beginning of the database recall, one year more for WEISKRANTZ (WEL), approximately ten years more for SPERRY (SRW) and much more for MORUZZI (MOG). So for author student-mentor pairs that are potentially adequately captured within the available timeframe, it is possible that there is a relationship between citing identity type and relationship to co-authors, in that younger researchers frequently will co-author with more experienced researchers and this may be reflected in citing identity types. This is

clearly also possibly interacting with the relationship between productivity, higher citing frequency and by extension higher citing identity type, combined with the likelihood for longer careers to have higher productivity.

Collaborations as they are specifically represented in the formal communication behaviours of participants are formed by a diverse range of individuals, diverse in experience levels and approaches, coming together over a specific ‘problem’. Citing identities are shown to be similarly diverse, a range of citing identities are captured within each collaborative grouping according to both MDS and cluster solutions. While co-authors clearly share publications, the nature of collaborative relationships as demonstrated within Consciousness may indicate that a range of individuals with diverse career and representational characteristics form these formal collaborative relationships, and this is aligned with a diversity of citation identities. However this also indicates that collaborative activity as defined by direct co-authorships and shared co-authors does not influence citing identities, but rather emphasises their differentiation.

5.4.3. COMMUNICATION CHANNEL SIMILIARITY BASED ON JOURNAL CO-PUBLICATION MATRIX ANALYSIS

To assess the similarity of Citation identity distributional types in terms of publication within similar journal groupings a combination of cluster analysis, multidimensional scaling and journal content title analysis was completed on an author-by-author symmetrical matrix derived from an asymmetrical author-journal matrix as outlined in Section 5.3.2.2.1.

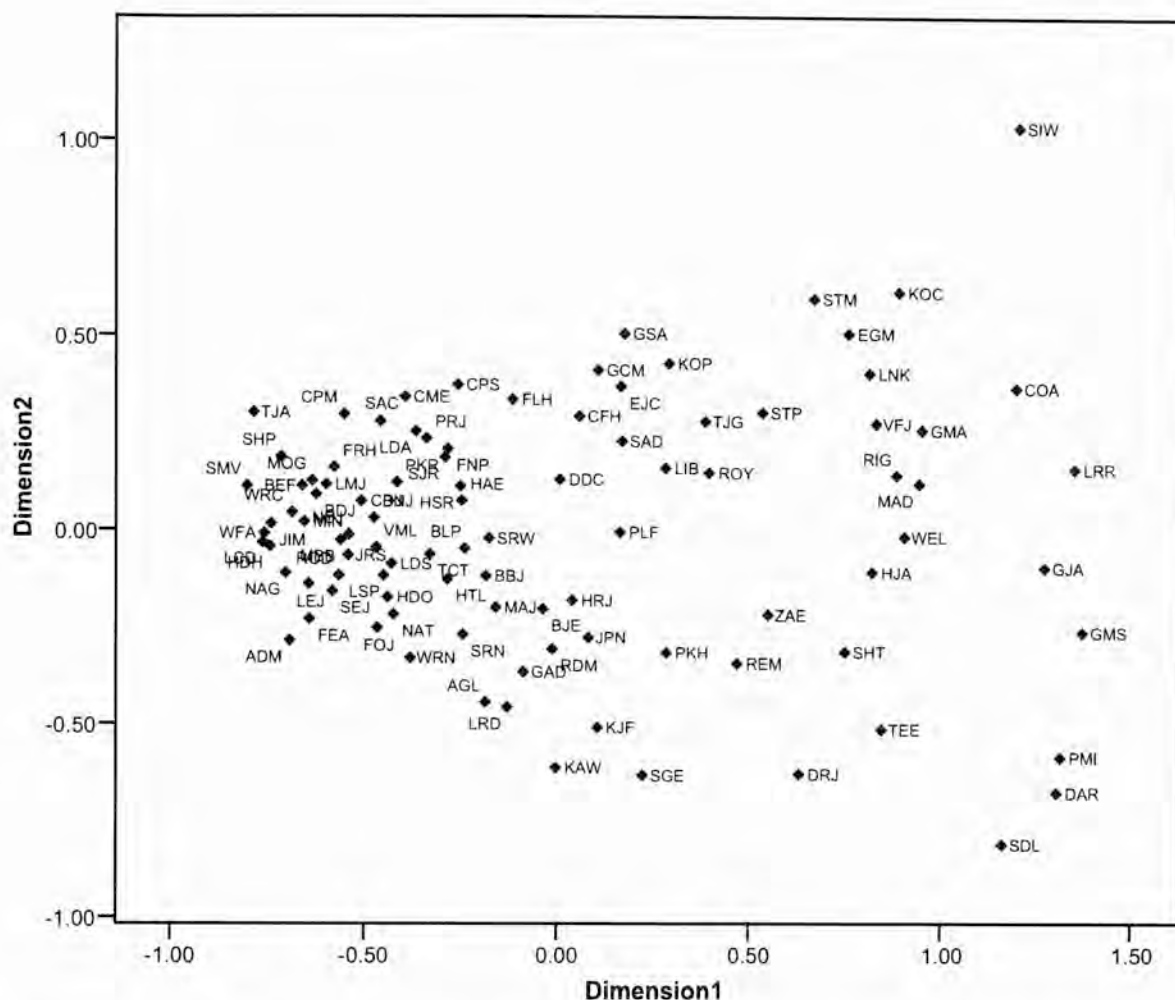


Figure 5.10 Euclidean distance multidimensional scaling solution for symmetrical author-author journal publication matrix labelled by author

The Multidimensional scaling (MDS) solution (using Euclidean distance) is presented in Figure 5.10. This solution provides a visualisation of the relationship between the authors according to their publication in journals across the Consciousness journal set. Authors that have published in the same journals are placed in closer proximity to one another than those that do not. The two dimensional solution was found to represent the data well with an acceptably low stress level of 0.02, high dispersion accounted for (0.98) and measure of fit (Tucker's coefficient of congruence = 0.99). The Euclidean solution in general conforms to similarity as assessed from the symmetrical raw matrix, for example SUSAN GREENFIELD (GSA) and ALFRED KASZNAK (KAW) who appear relatively distanced on Dimension 2, have a low similarity measure of 182.38 while KASZNAK (KAW) and SCHWARTZ (SGE) have a much higher similarity measure of 298.01. However the MDS solution is not entirely consistent as demonstrated by comparison of DAMASIO (DAR) and GREENFIELD (GSA) who also appear at a relatively large distance from one another in the MDS solution but have comparatively high similarity according to the alternate measure (260.98). This is due to the optimisation

function of MDS where interactions with other variables will influence the distances between pairs of individual authors. MDS here is combined with other methods of analysis to attempt to overcome some of these limitations. In general Dimension 2 gives a broad indication of similarity between authors according to their journal publication activity, which will be further analysed using cluster analysis and content analysis below.

Dimension 1 however is found to have a strong statistically significant relationship to the productivity of individual authors within the Consciousness authors' journal set.

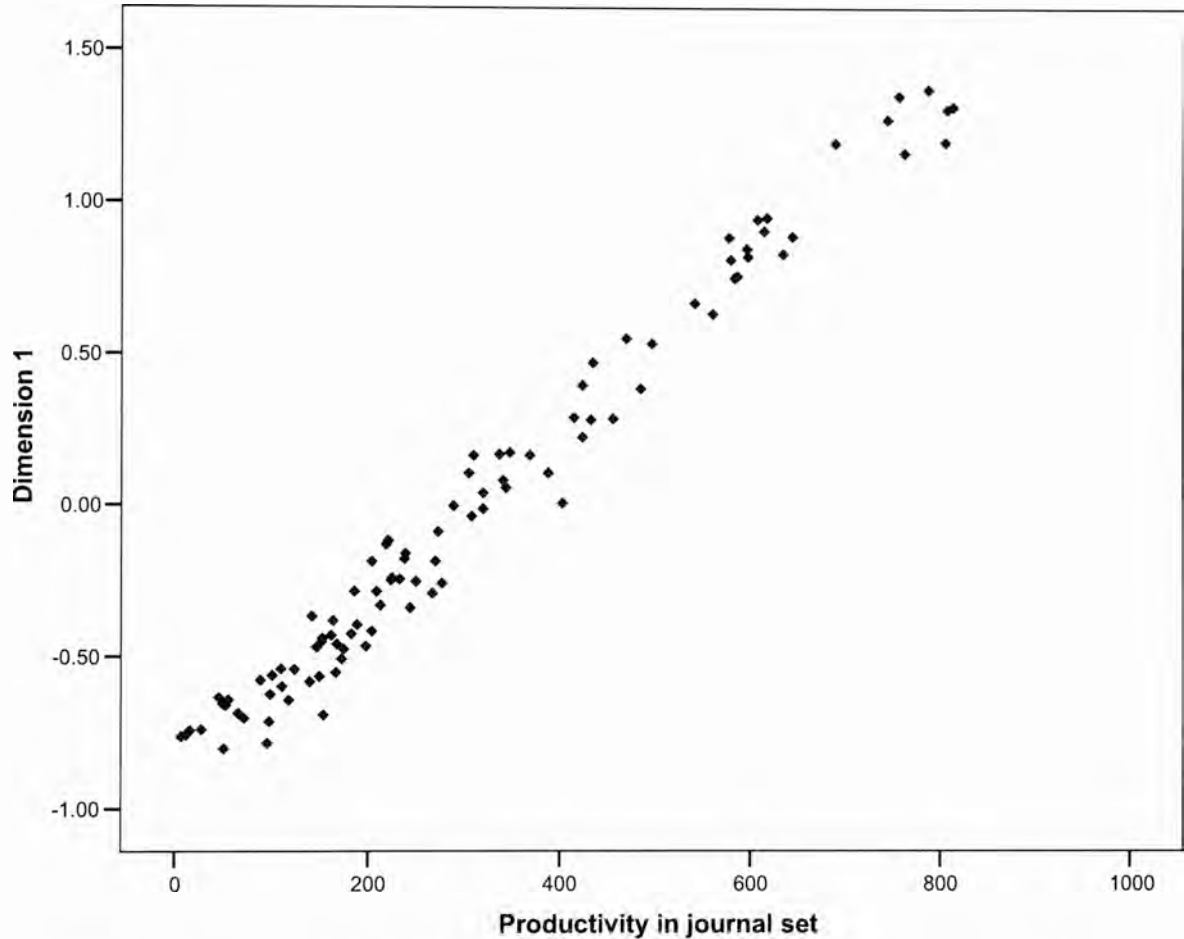


Figure 5.11 Linear regression of author-author productivity within Consciousness journal set

Figure 5.11 demonstrates a clear relationship between productivity in the shared Consciousness journal set (individual authors' productivity in the journals within the common journal set published in by consciousness authors) and Dimension 1. This means that Dimension 1 indicates a strong relationship to the amount an author publishes within the journals published in by all the authors in the Consciousness journal set. This relationship is very strong at $R^2= 0.98$. The two variables were found to correspond to a normal distribution according to a one-sample Kolmogorov-Smirnov

test (both Asymp. Sig. (2-tailed) were greater than 0.05, Dimension 1 = 0.60 and productivity in journal set=0.68). The residuals are also normally distributed confirming appropriate use of linear regression.

A cluster analysis (using Ward's method) was then applied to the author-author symmetrical matrix to further examine appropriate journal content groupings relationships between the authors. A total of 7 groupings, when overlaid on the MDS analysis, were found to correspond reasonably well (Figure 5.12).

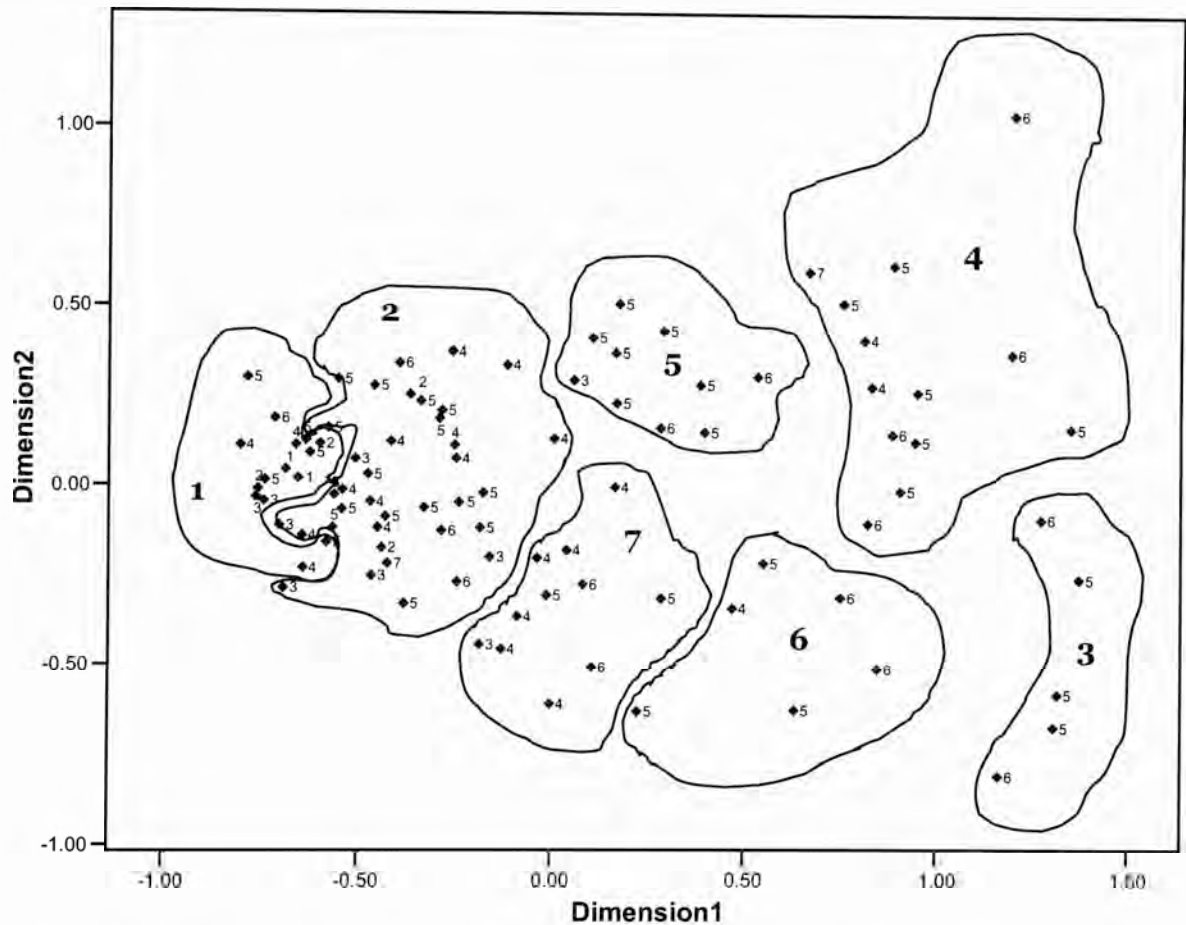


Figure 5.12 MDS Euclidean distance solution labelled with citing identity types and overlaid with journal content cluster Ward's solution

To determine the underlying content relationships of these journal clusters a content analysis was performed on the titles of the journals that correspond to the clusters. A considerable amount of overlap of journals used by authors within the cluster was found (Table 5-2), the diagonal contains the number of journals published in by authors unique to that cluster.

Table 5-2 Number of journals common between the clusters

	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6	cluster 7
cluster 1	77						
cluster 2	57	207					
cluster 3	18	88	47				
cluster 4	27	116	119	114			
cluster 5	33	92	70	113	47		
cluster 6	14	73	105	82	46	44	
cluster 7	27	101	116	87	56	94	83

The overlap evident in Table 5-2 indicates that the clusters are highly inter-related and that authors frequently publish across the journal clusters within the set, diminishing the extent to which content boundaries can be identified according to journal publication activity. This is in keeping with the possibility of broad scope of content captured within the journal selection and publication process.

Overlap in journals published in by the journal clusters were the more generalist publications or those specialised on the topic of Consciousness (refer to Appendix D).

In order to introduce differentiation between cluster groupings, a journal title content analysis was performed on journals that are unique to each individual cluster. This analysis found that journal cluster 1 and 2 were more highly differentiated by their use of their largest topic category grouping, so publication in journals with these keywords in their titles, the other five clusters had a greater spread of topics across their categories.

Table 5-3 Journal title content clusters for journals unique to journal cluster

cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6	cluster 7
Physics (35) (40% total)	Philosophy (111) (41% total)	Neuroscience Psychology (7 each) (14% total each)	Physiology (25) (23% total)	Bioscience (10) (23% total)	Medicine (11) (26% total)	Psychology (19) (23% total)

Clusters 1 and 2 can be seen to have more highly differentiated content groupings according to the journals in which they are publishing. However even when considering only the journals that have been uniquely published in within each journal cluster, there is a considerable amount of interconnection. While unique journal publication activity in cluster 2 is predominantly focused in journals with philosophy in their title this is also the largest group, having both the largest number of authors encompassed and correspondingly the largest journal set.

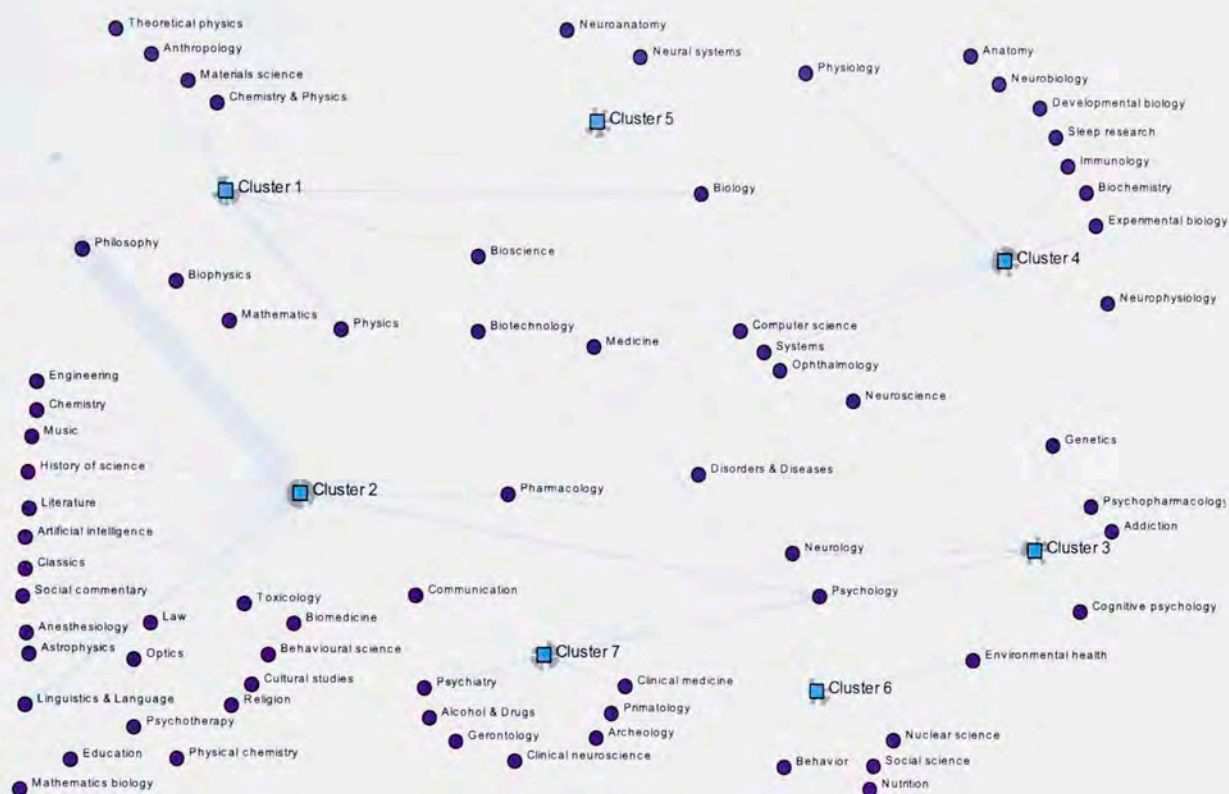


Figure 5.13 Network representation of connections between journal title keyword topic categories, according to unique journal usage by journal clusters⁵⁸

Figure 5.13 demonstrates both the interconnectivity of the topic categories of the journals used uniquely by authors across the Consciousness journal set, and also that there are a number of differentiable content orientations within these journal clusters. Dominant categories for the journals cluster are evident in the network representation, for example cluster 2 is clearly most strongly connected to the PHILOSOPHY category and cluster 1 to the PHYSICS category (as according to Table 5-3). However the network representation puts these categories into further context by demonstrating the overlap apparent in the categories across the clusters. While cluster 2 is most associated with the PHILOSOPHY category cluster 1 also publishes in journals that include this keyword in their title. Cluster 2 and 1 are further connected via publication in PHYSICS, MATHEMATICS and BIOPHYSICS categories. Cluster 2 and 7 are connected via the COMMUNICATION category but also more dominantly by the PSYCHOLOGY category, to which cluster 3 is also connected but with a lesser degree of strength. The MEDICINE category is shown to be central to all the clusters; however it has its most dominant connection to cluster 6. Several categories form intersections between clusters, including BIOTECHNOLOGY which sits across clusters 1, 2, 3 and 4, and NEUROSCIENCE

⁵⁸ This visualisation only includes categories that are published in 2 to or more times for readability, the analysis described considers the entire matrix and so includes all categories used across unique journal set. The connecting lines' thickness indicates the frequency of publication within that journal topic category by that journal cluster.

between clusters 3, 4, 5 and 7. The BIOLOGY category also connects all the journal clusters. These category interconnections further define the nature of the content encompassed by the journal clusters, demonstrating that publication activity is across subject content, forming a combination of different approaches across delimitations of journal publications boundaries.

The differentiation seen in Figure 5.13 is in keeping with the use of journals across the complete Consciousness journal set. For each individual cluster the topic categorisation indicated by the analysis on the unique journals is consistent with the dominant publication in these groups by the clusters journal set. Selecting only the journals unique to each cluster to analyse has removed the more generalist publications such as Nature, though the representation continues to indicate the dominant content areas published within that cluster. For example, cluster 1 most frequently publishes in a PHYSICS journal, cluster 2 overall is dominated by PHILOSOPHY journals though does also publish in a large number of PSYCHOLOGICAL or NEUROPSYCHOLOGICAL journals, cluster 3 is dominated by BIOLOGICAL and COGNITIVE NEUROSCIENCE journals, cluster 4 by EXPERIMENTAL BRAIN RESEARCH, BEHAVIOURAL AND BRAIN RESEARCH and NEUROSCIENCE, cluster 5 in terms of the journals dominantly published in is follows closely cluster 4 however closely followed by NEUROPHYSIOLOGY and PSYCHOPHYSIOLOGY journals, cluster 6 is even less differentiated when considering the total journal set published in by this cluster, MEDICINE does appear frequently but this cluster more frequently publishes in COGNITIVE NEUROSCIENCE journals and neuropsychological journals, cluster 7 is again more clearly dominated by PSYCHOLOGICAL journals though there is also a large representation of NEUROPSYCHOLOGICAL and BEHAVIOURAL BRAIN journals.

Differentiation shown in this content analysis is not particularly surprising given that journals were selected according to their unique publication use by each cluster, though this is also supported when the full set for each cluster journal set is considered. However what is more surprising is the interconnectivity demonstrated within the topic categorisations. This interconnectivity does highlight the limitations inherent in the specific definition of scientific content via journal groupings; however examining journals does begin to capture potential differentiation according to broader categories of journal content scope, and on the individual journal level potential differentiating in citing behaviours related to journal publication activity, with possible implications for editorial influence or broad disciplinary based citing practices.

5.4.4. COMMUNICATION CHANNEL SIMILIARITY AND CITING IDENTITY TYPE

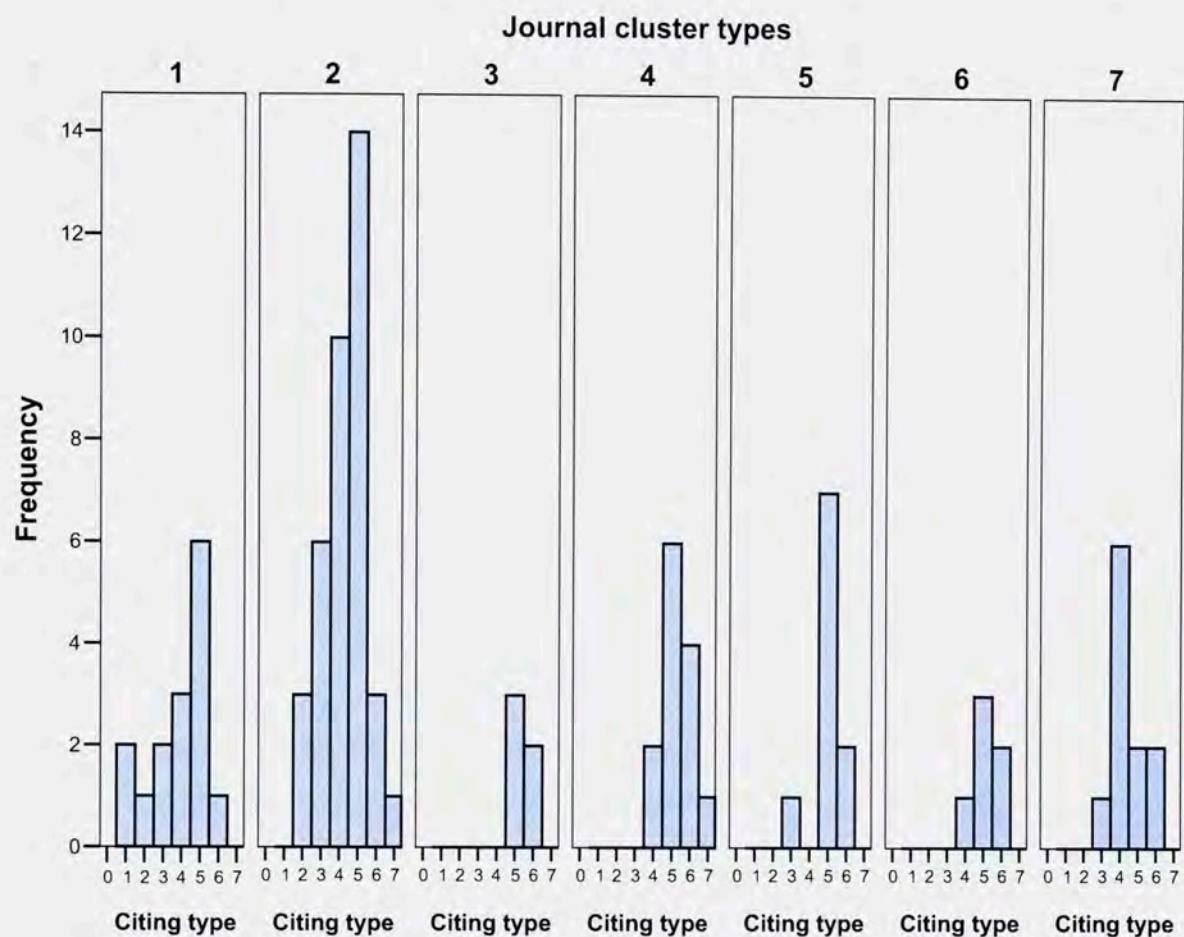


Figure 5.14 Citing identity types and their frequency of appearance in journal clusters

Figure 5.14 shows the frequency of occurrence of each of the citing identity types according to their placement in journal clusters (and thereby the authors' similarity of co-publication within the consciousness journal set). Journal cluster 2 identifies authors that publish most frequently within a journal sub-set that can be identified with the dominance of the title keyword PHILOSOPHY. This cluster also has the most authors identified as belonging within it; accordingly it is not surprising to find that all but one citing identity type 1 is found within this journal cluster. There is therefore a large range of diversity of citing distributions captured within this journal cluster. Similarly, journal cluster 1 encompasses a range of citing identity types excluding only citing identity type 7. However, notably citing identity type 1 only appears in journal cluster 1; citing identity type 2 appears in journal clusters 1 and 2 only, with citing identity type 2 being dominant in this cluster; citing identity type 3 appears in journal clusters 1, 2, 5 and 7; whereas citing identity type 4 appears across journal clusters 1, 2, 4, 6 and 7, though most frequently in citing identity type 2. Citing identity types 5 and 6 have representatives in all journal content clusters, whereas other less populated

groupings also have less representation in diversity of clusters. However, the higher journal clusters can be seen to be dominated by the higher citing identity types, with the exception being citing identity 7 which appears in the journal cluster types 2 and 4, which is significantly diverse given there are only two authors within this citing identity type. This approximate trend is worthy of a closer inspection.

Journal clusters 1 (PHYSICS) and 2 (PHILOSOPHY) sit between -1.00 and 0.00 approximately on the Dimension 1 axis in the journal co-publication MDS solution (Figure 5.12). Journal clusters 7 (PSYCHOLOGY) and 5 (BIOSCIENCE) occur approximately in the same space according to dimension 1, between 0.00 and 0.50. Journal clusters 6 (MEDICINE), 4 (PHYSIOLOGY) and 3 (NEUROSCIENCE) all occur at the far end of dimension 1. Cluster 6 occurs between 0.50 and 1.00, cluster 4 beginning at approximately 0.70 and ending at 1.40. Finally cluster 3 is wholly contained between 1.00 and 1.50 on Dimension 1.

It was found earlier in this section that Dimension 1 was strongly and statistically significantly correlated to publication productivity within the journal set, so therefore progression of these journal clusters along Dimension 1 implies an increase of productivity through the clusters, indicating some form of productivity relationship to the content similarity represented by the cluster groupings. In order to statistically test implications of productivity and content across Dimension 1 the Gini citing distribution measure is used (as it is strongly related to the citee-citation ratio it can be used to represent the relation, and for the most discriminating linear regression possible here an interval measure is preferred, as opposed to a nominal measure as produced with clustering). When Dimension 1 is regressed on Gini only a very weak statistical correlation ($R^2=0.25$) is found; however the correlation is significant and the residuals are normally distributed. This implies, though only very weakly, that content has a relationship to productivity and that productivity has a weak relationship to citing concentration and citing identity types. This indicates that further discriminating content analysis and journal publication activity is warranted.

Dimension 2 on the other hand has no relationship to Gini at all, which is visually apparent in Figure 5.12, where citing identity types can be seen to be scattered throughout Dimension 2. The content relationships expressed across Dimension 2 are less apparent. Dimension 2 does indicate approximate similarity within journal publication activity; however clear discrimination of the basis of this similarity is elusive. Clusters 5, 7, 6, 3, 4 do form into coherent groups across Dimension 2; however these groups have a considerable amount of overlap in journal content and publication

activity, therefore not allowing enough discrimination between content and journal publication activity to demonstrate convincingly whether citing distribution according to the citing identity relation is related to this dimension of content.

5.4.5. SELF-PLACEMENT OF PARTICIPANTS IN REPRESENTATIONAL SPACE - CITEE MATRIX ANALYSES COMPARED TO CITING IDENTITY TYPES

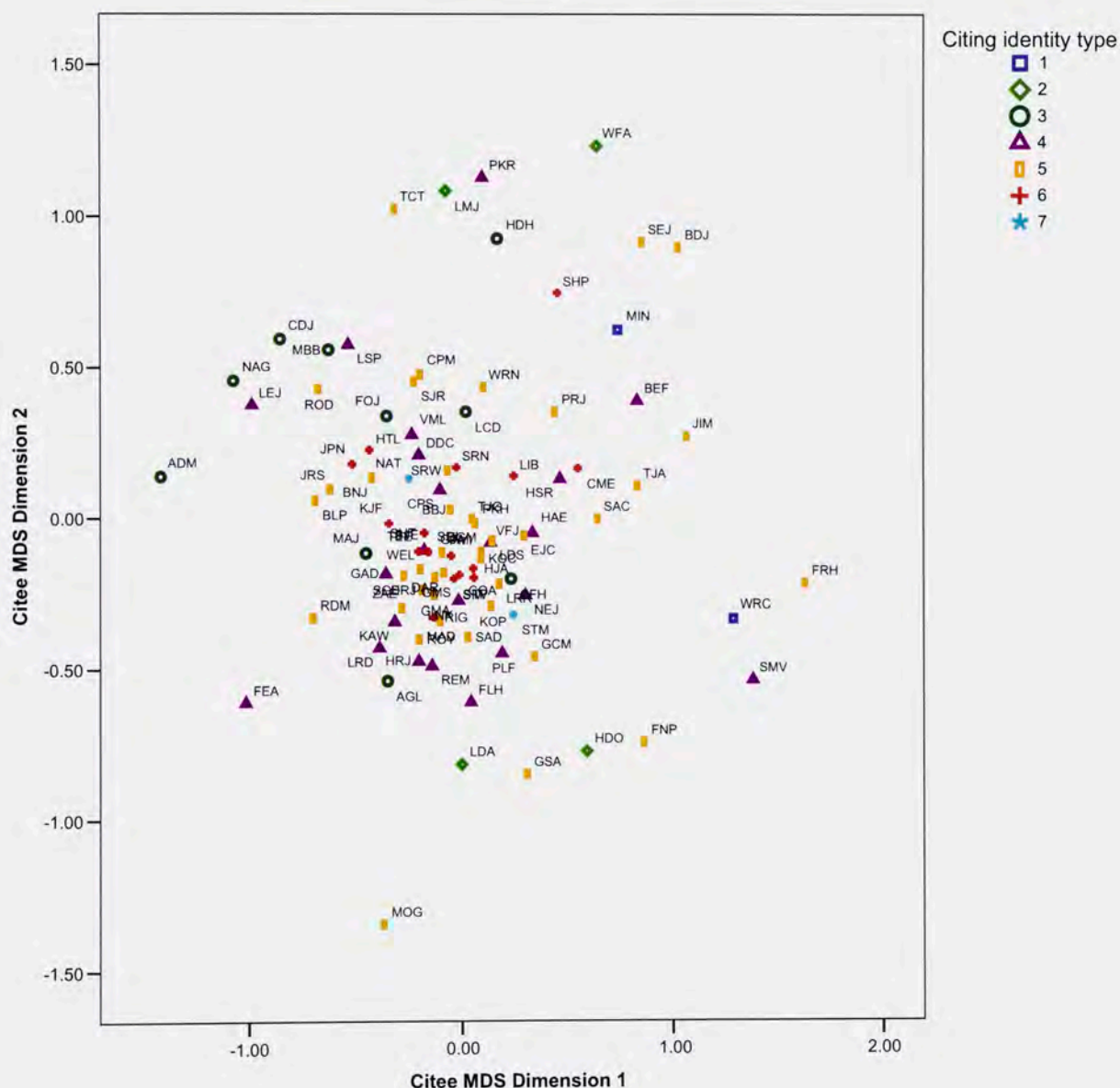


Figure 5.15 MDS of citee similarity matrix according to citing identity type

Figure 5.15 is an MDS solution developed from a citee-similarity matrix, as outlined in the methodology section 5.3.2.4. This MDS is a visualisation of participants co-citing patterns; the matrix is comprised of the collection of all citing identities (original asymmetrical matrix) which was then re-calculated as to how often each participant shared a citee (symmetrical similarity matrix). It was expected that if any of the

representational spaces presented in this analysis, this would be the one most likely to produce possible content similarity in citing patterns.

Normalised raw stress for this solution is appropriately low at 0.04 indicating that the misfit of solution to the data is low while dispersion accounted for is relatively high at 0.95, and Tucker's coefficient of congruence measure the fit is also good at 0.98. This further supports the applicability of this solution, in that it provides a good representation (with limitations discussed below) of the underlying similarity matrix, is in accord with recommendations for use of this MDS methodology by Leydesdorff and Vaughan (2005 preprint).

A citee cluster analysis was performed resulting in a 10 cluster solution. This solution mapped reasonably well to the MDS presented above; however contained a significant portion of undesirable chaining, therefore 3 clusters had only one author, a further cluster had only 2, and 3 more clusters had only 3 participants. This 'best' solution is therefore not adequately informative as to the grouping of authors according to their citee content groupings, and is not presented as a solution.

Even though the same data is used for the co-citee similarity matrix and citing identity types, there are no significant patterns of correlation in evidence (Figure 5.15). This method is clearly not sufficient to further understand citing identities and their content relationships to citees in the Consciousness data set.

5.5. RESULTS - CITING IDENTITY TYPE AND CITING BEHAVIOUR ANALYSES

5.5.1. PRODUCTIVITY

Based on the calculation of the Gini, authors with high productivity will have a tendency to have a high Gini value. This is due to an expectation of a high level of re-citing, whereas authors with low productivity will have a tendency to have a higher proportion of unicites. A viewing of the data in Figure 5.16 demonstrates this in the main to be the case. There are only low producers that also have a low Gini value, while authors with a high Gini value can be high or low producers. Substituting simple productivity with fractional productivity⁵⁹ made little difference to the outcomes.

⁵⁹ Fractional productivity was calculated by attributing a weighted productivity based on the number of authors per publication.

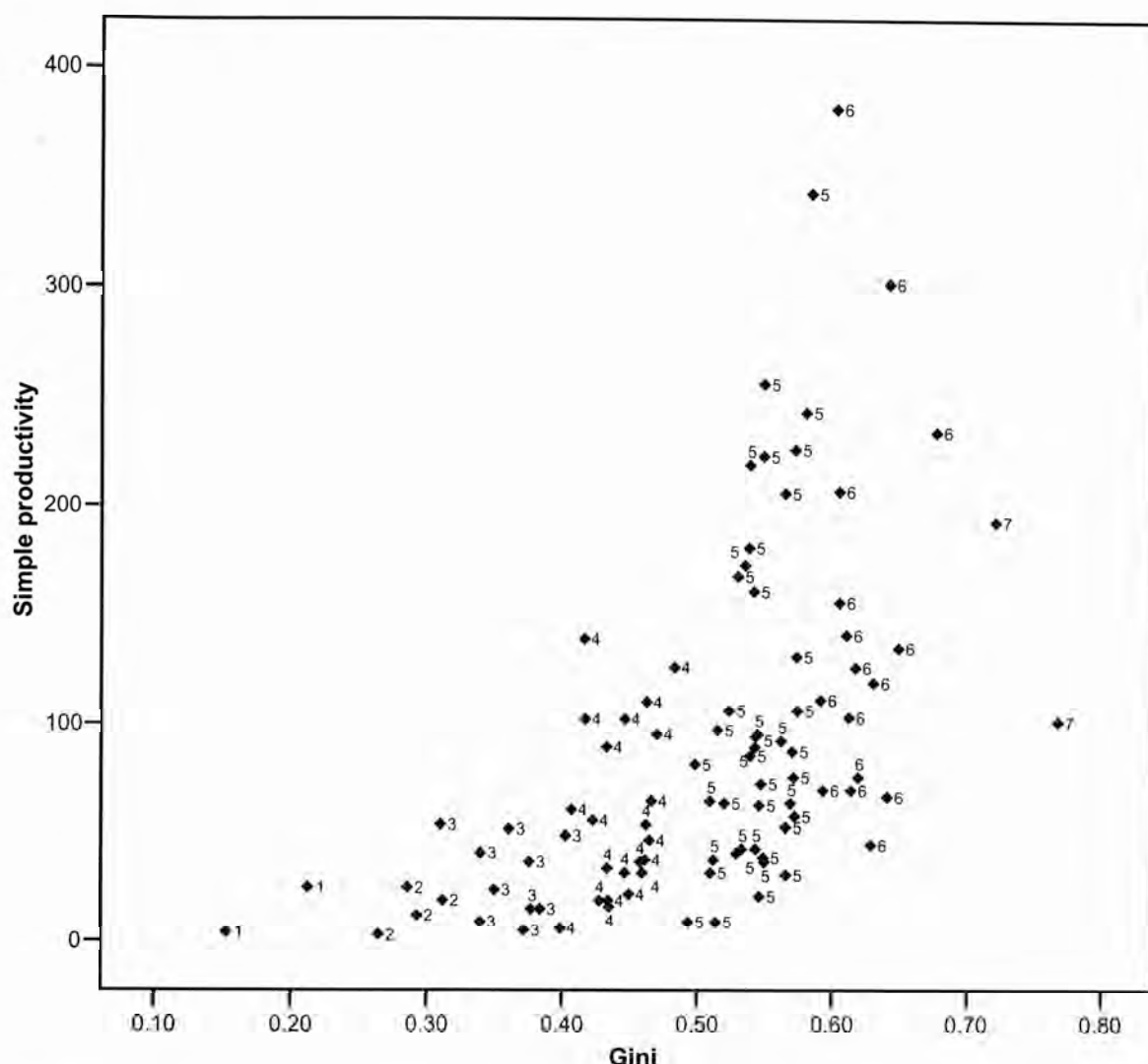


Figure 5.16 Simple productivity related to Gini measure of citing density by citing identity types

5.5.2. LENGTH OF PUBLICATION CAREER

Length of publication career was taken to be the first publication date available until 2001 (or year of last publication of author available) from the searches described in Section 4.2.4. Publication career length is shown by citing identity type for the Consciousness authors in Figure 5.17. There is little correlation between publication career length and citing identity type. It is apparent that citing identity types 1 and 2 have no researchers of more than 25 years publication career length, from either being a relatively young researcher or retired/deceased.⁶⁰ However the restriction of publication career length may be due to the limited sample of participants. As expected, an extended publication career can result in high levels of productivity; however many authors in the Consciousness set display relatively low productivity levels despite

⁶⁰ See Appendix D.4 for publication career length author details

having long careers. A graph of simple productivity and career length is shown Figure 5.18.

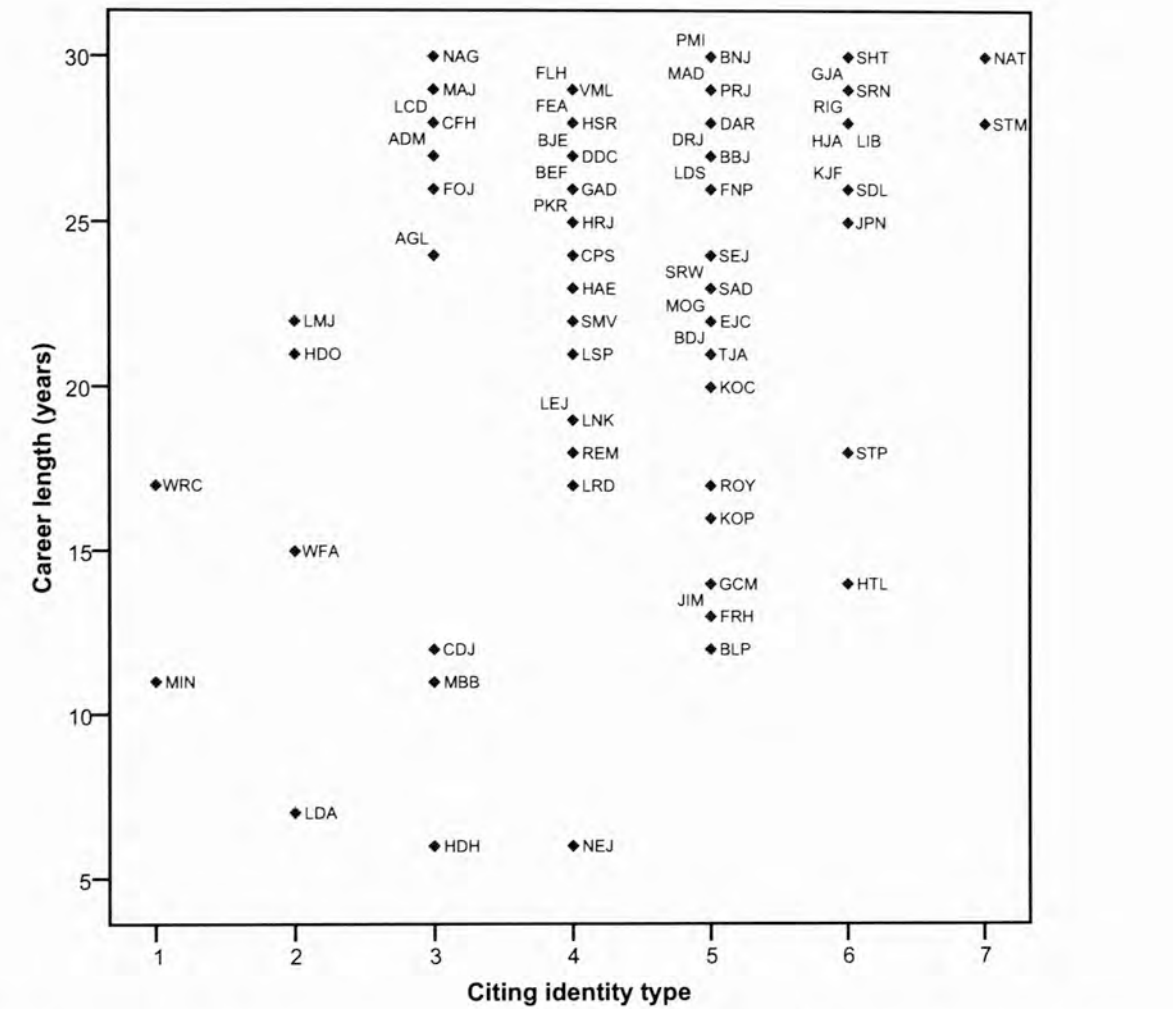


Figure 5.17 Length of publication career as represented in ISI by citing identity type

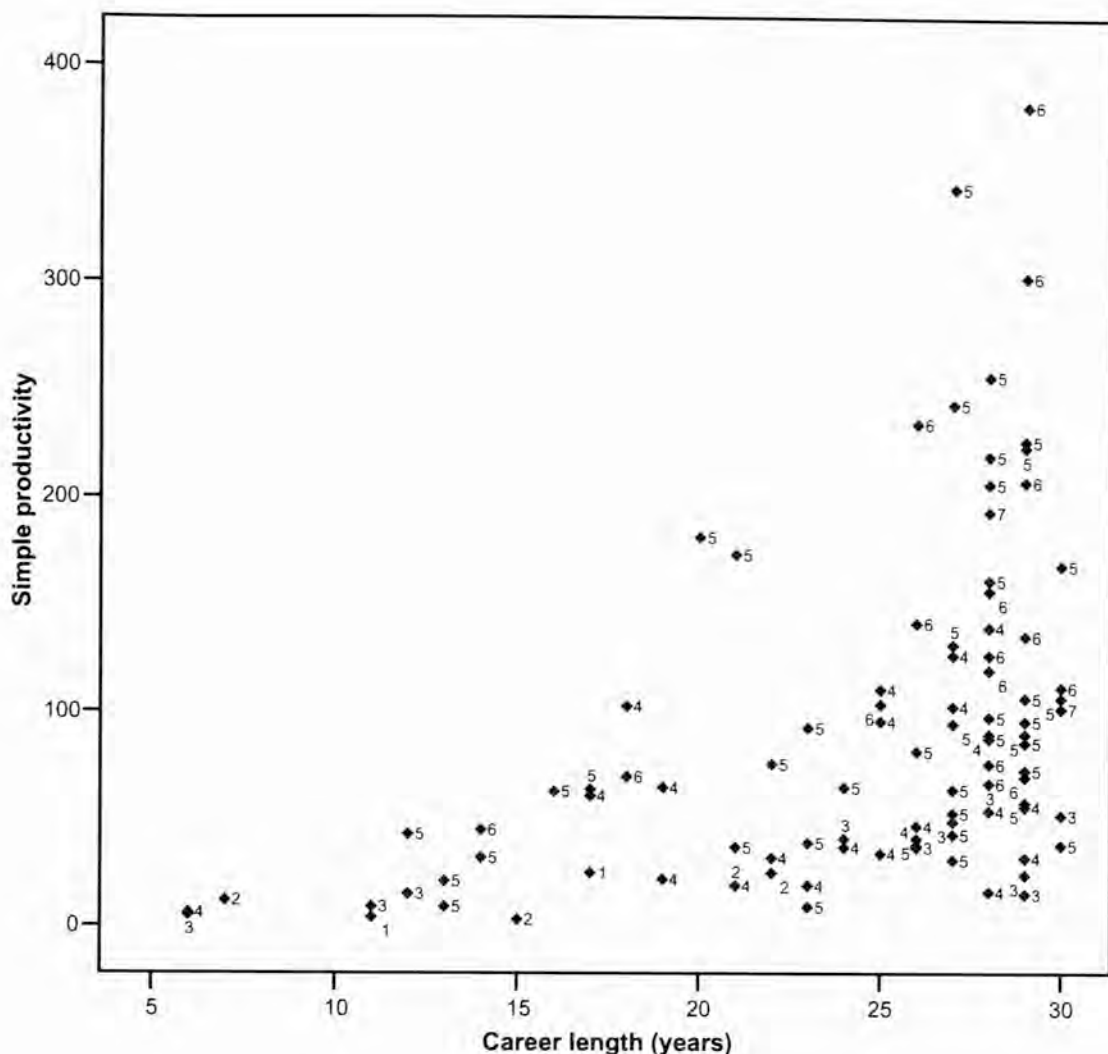


Figure 5.18 Simple productivity in relation to career length in years, by citing identity type

5.5.3. UNICATION (UNICITES)/RECITATION (RECITES)

White (2001b) introduces two complementary measures into his analysis of citation identities: percentage of unication and percentage of recitation. The concepts to which they refer shall be here called unicites and recites respectively, in order to maintain internal constancy. These measures essentially are introduced by White to further capture an individual author's citing distribution and to amplify his citee-citation ratio analysis. The total measures, an individual's total unicites or recites, taken separately have no clear relationship to the Gini or the citee-cites ratio; however when the measures are averaged over the authors total cites, the Gini captures a high amount of their explanatory variance; $R^2=0.90$ for both unicites over total cites and recites over total cites, the first being negatively correlated and the second positively. A chart of normalised unicites versus citing identity type is given in Figure 5.19, which clearly shows this trend. This is not a particularly surprising result as the Gini addresses the distribution the two measures represent, though it does suggest that the Gini does so

comprehensively. In terms of the citing identity types developed for this analysis, the unicate and recite measures are clearly seen to be an element contributing into an individual's Gini versus citee-cites ratio relation. While an example from Section 4.4 showed how COWEY (COA) and HOBSON (HJA) had dissimilar unicate levels but equivalent Gini values, this is an exception rather than the rule.

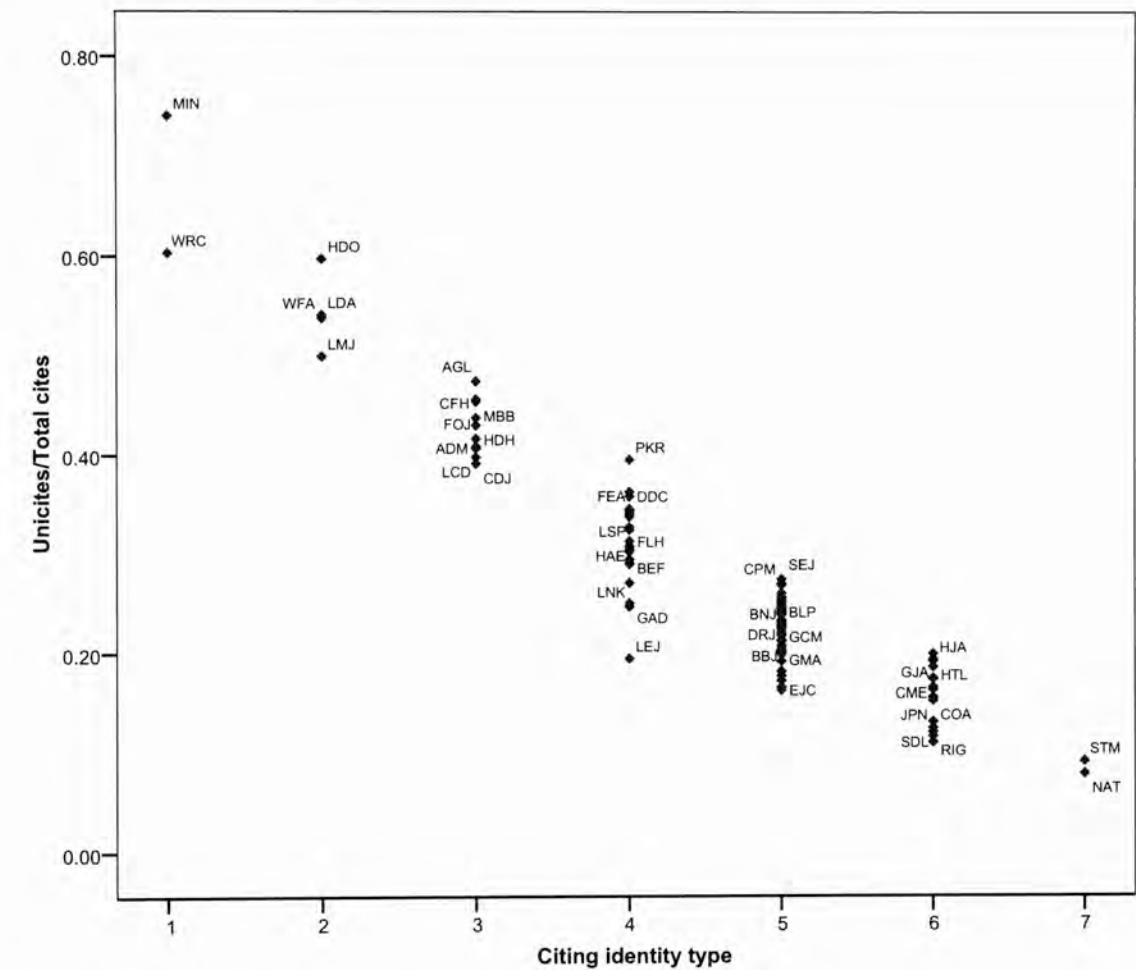


Figure 5.19 Average of unicites according to citing identity types

When citing clusters are examined using the next level of concentration, the normalised recites only, the trend continues. There is a progression up the Gini versus citee-cites relation from cluster 1 through to cluster 7.

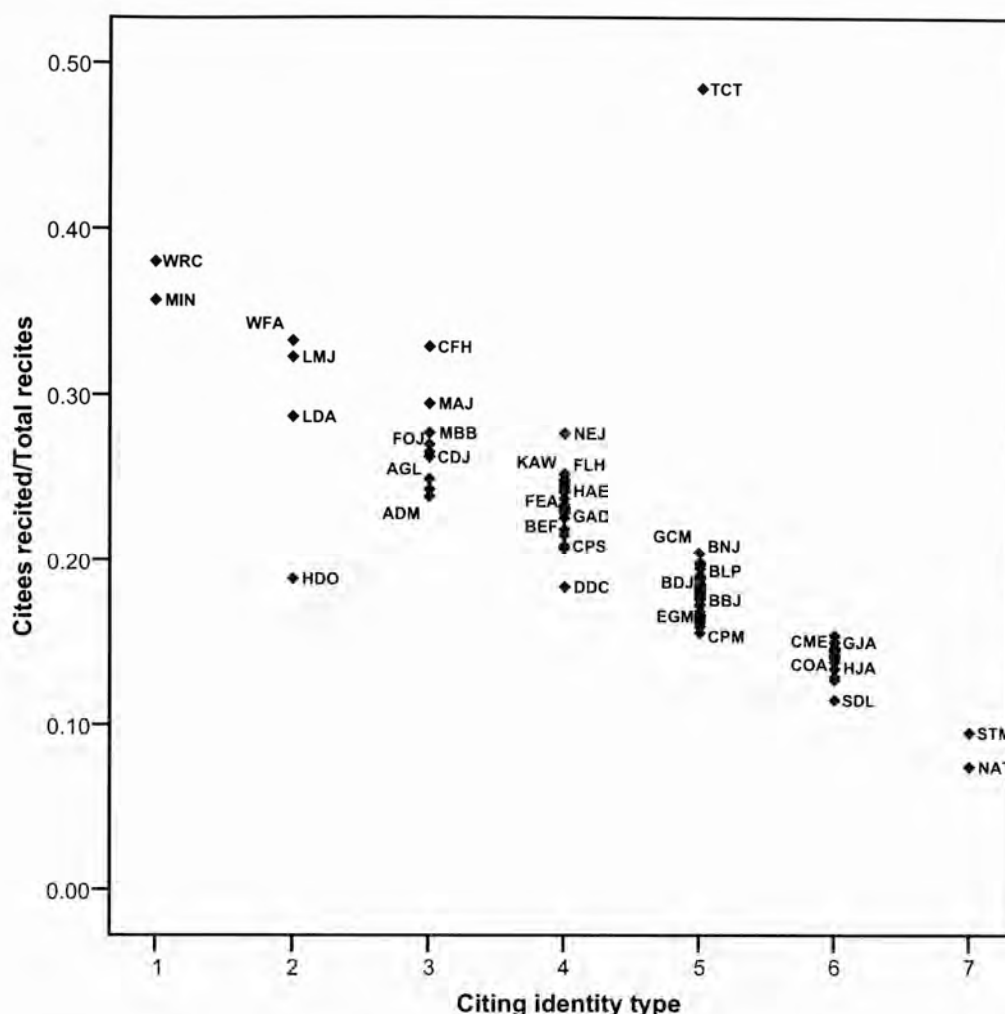


Figure 5.20 Reciting (citee recited/total recites) patterns according to citing identity types

As can be seen in Figure 5.20, clusters 7, 6 and 5 are generally placed lower, and so tend to use their recitees more in relation to their recites, whereas authors in clusters 1, 2, and 3 tend to be more sparing in their use of their recited authors. Again cluster 4 lies between the two more extreme clusters. There are two individuals that are exceptions to this general progression: CHARLES TART (TCT) (placed some distance above the majority of cluster 5) and DONALD HEBB (HDO) (placed somewhat below the general group in cluster 2). Both CHARLES TART and DONALD HEBB have relatively high self-cites percentages, TART has the highest and HEBB the seventh highest self-cites in relation to the rest of the set, this condition is combined with a significant drop in cites to other recites. TART cites himself 237 times but the next most frequently cited citee jumps down to only 14 cites, his average of cites to recitees not including himself is only 3.7. HEBB has a similar though not so extreme pattern, where he cites himself 22 times and the next most highly cited citee is only cited 3 times. Perhaps significant to the observation of the patterns evident in DONALD HEBB's unusual re-citing patterns is that this analysis only captures the late stage of his publication career (Appendix D.4). HEBB

completed his PhD in 1936, and published the majority of his significant work prior to the beginning of the scope of the ISI citation indexes which limit the time frames for this analysis. Within the ISI citation data set collected for this analysis, HEBB's first publication is in 1973, within the Science Citation Index and his last was published in 1994: this is clearly a tribute piece as HEBB himself died in 1985. The other publications of HEBB that were captured in this analysis are within his lifetime and do represent some of his last publications, and so the recite-recitation rate identified within Figure 5.20 appears to be capturing this to some degree.

5.5.4. SELF-CITING

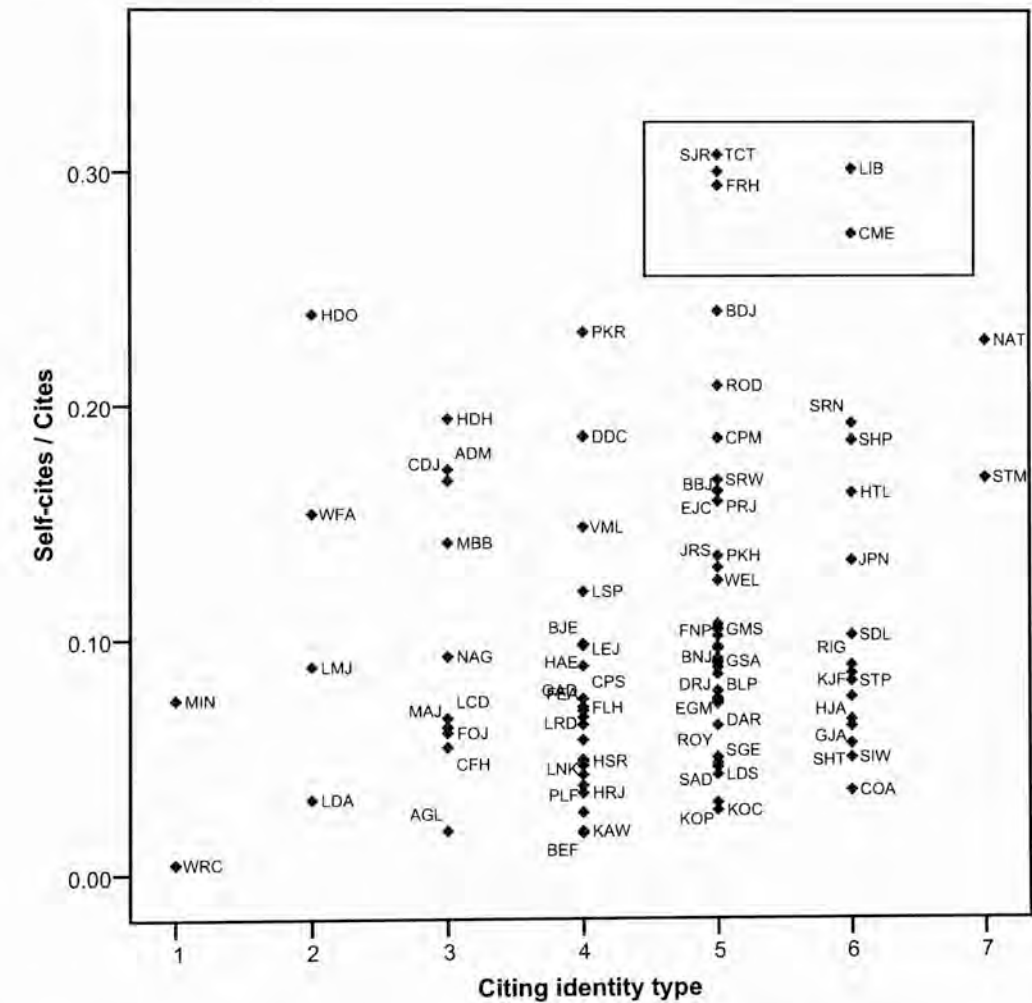


Figure 5.21 Self-cites normalised by total cite yield by citing identity type

Figure 5.21 shows that self-citing by authors in citing clusters 2 through 6 are equally distributed; no clear differentiation is evident between self-citing patterns across these types. Differentiation can be seen between type 1 and type 7; both authors in type 7 cite themselves more frequently than the two authors in type 1. However self-citing in type 7 is not clearly higher than that found in type 6 and therefore there is no direct

relationship between self-citing activity and placement according to the relation. By extension self-citing cannot be seen as a straightforward contributor to the definition of individual authors citing distributions.

Contrary to expectations, rate and percentage of self-citing by authors appears to have no significant role in contributing to their placement along the relation, and therefore their citing identity type, nor within their overall distribution as determined by the Gini measure alone. Unlike the pattern established for unicites, no equivalent pattern is discerned for self-citing behaviour, which is generally the highest cited author.

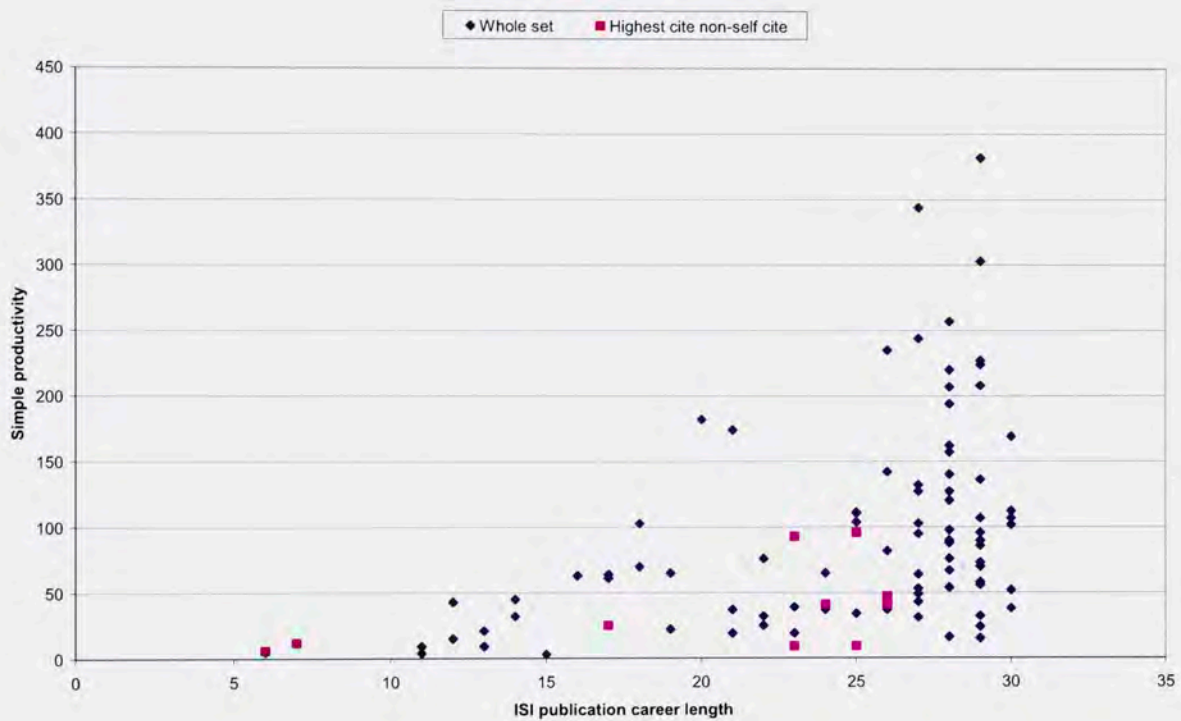


Figure 5.22 Authors that do not self-cite most frequently compared with authors that do, according to simple productivity counts and ISI publication career length.

10 of the 98 authors do not cite themselves most frequently. Of these two cite themselves an equal amount to another author. Essentially this occurs only with low productivity authors, but as can be seen in Figure 5.22 there are exceptions. It was not expected that authors with greater than 10 years career length would cite others more highly than themselves.

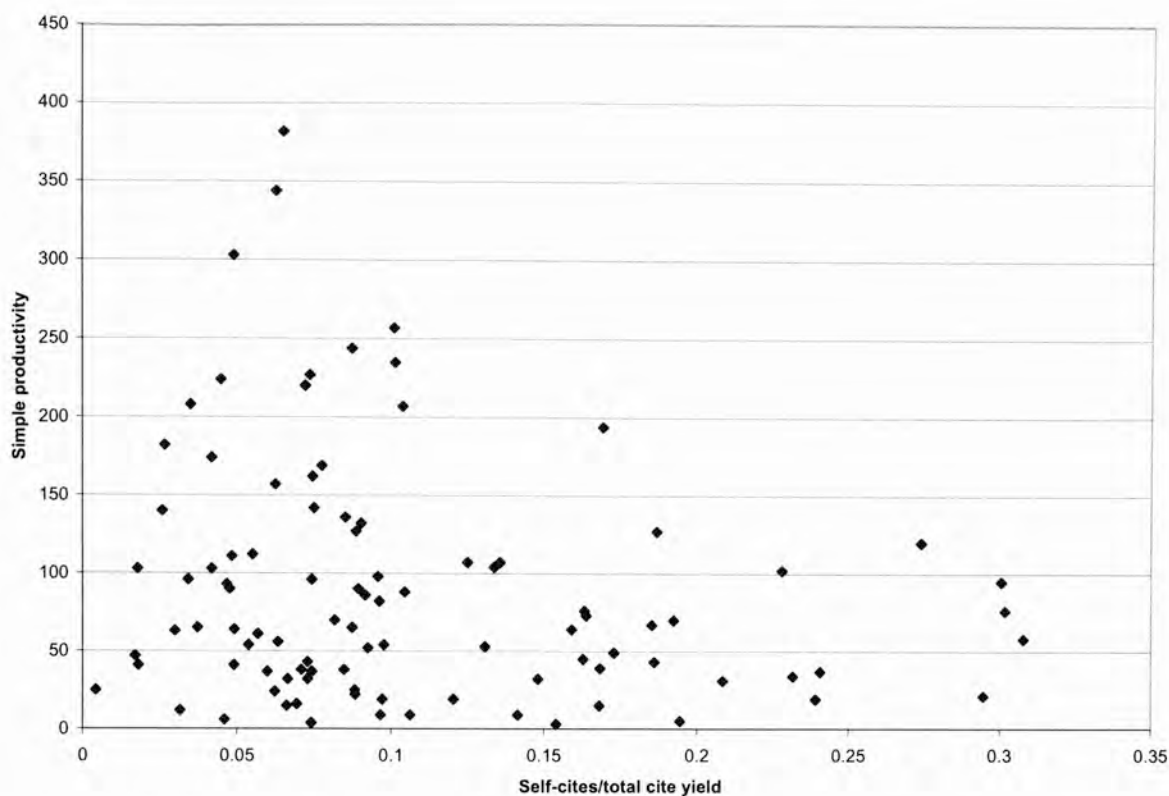


Figure 5.23 Simple productivity in relation to self-citing activity normalised for total cite yield

Increases in productivity are not accompanied by increases in the proportion of self-citing (Figure 5.23). The five authors that fall around the 0.3 range of self-citing all have productivity rates at lower than 150 publications each, whereas authors with higher productivity rates of 300 to 400 publications tend to have lower self-citing rates of between 0.05 and 0.07.

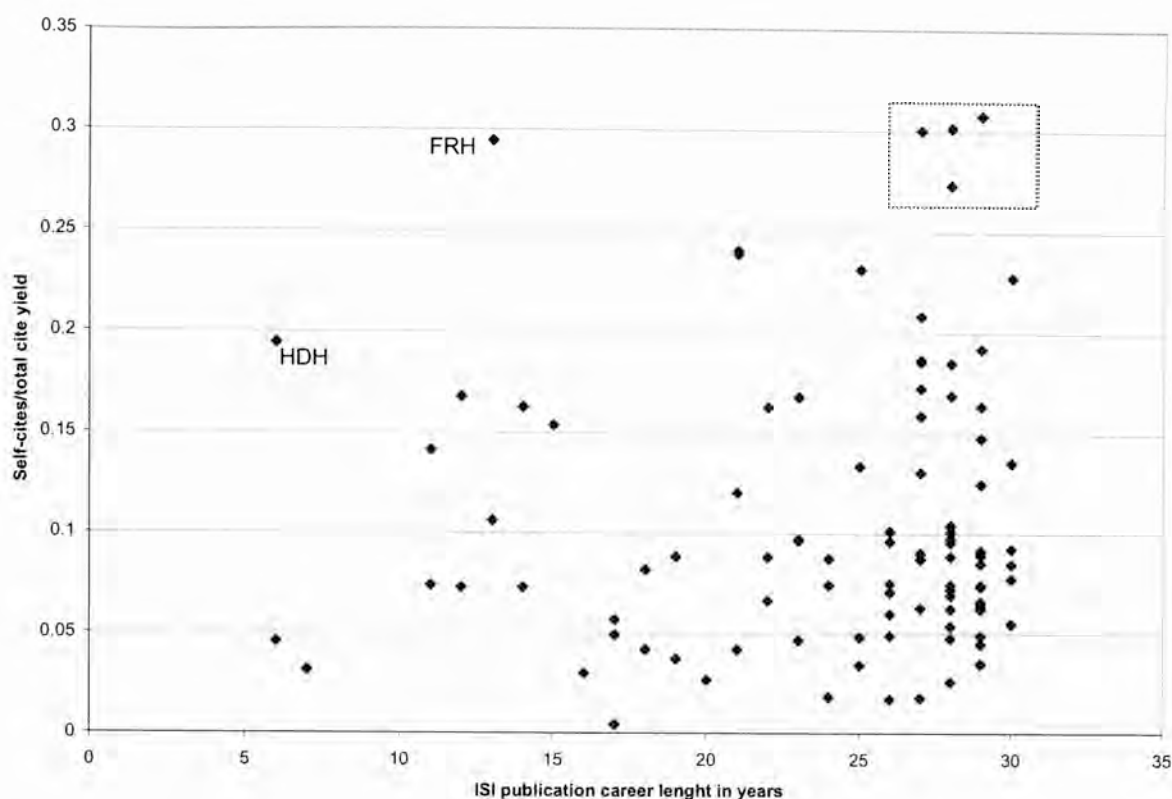


Figure 5.24 Self-citing patterns normalised for cite yield by ISI publication career length in years

Figure 5.24 indicates that for the majority of authors self-citing increases as their publication career lengthens. However, as can be seen the amount of self-citing activity in relation to total cite yield varies widely even among authors that have similar length careers. The dashed box in the top left corner corresponds to the authors appearing in Figure 5.21, SEARLE (SJR), TART (TCT), CONRAD (CME) and LIBET (LIB). These authors maintain a relatively high rate of self-citing for both their citing type and in relation to their colleagues with similar lengths of publication career. The exception to this is HERBERT FROHLICH (FRH). FROHLICH presents as having an unusually high self-citing rate for an author with a publication career length of 13 years; however this analysis captures only the late stages of FROHLICH's publication career. HERBERT FROHLICH was born in 1901 and died in 1991, receiving his doctorate in 1929. Even so FROHLICH has a high self-citing rate but given the length of his publication career it is more in keeping with a number of his contemporaries. DAVID HODGSON (HDH) is another author that is seen to have a relatively high self-citing rate in relation to his relatively short publication career length. He also is a particular case in terms of the career path that underlies this publication activity. In comparison to other authors with 6 year publication careers, HODGSON received his doctorate relatively early in 1965 at the University of Oxford; however he did not pursue a research career, but rather progressed to become a Supreme Court Judge at the Court of Appeals, Supreme Court

of New South Wales, and in 1994 he transferred this position to be at a part-time level. It is in 1996 that his publications begin to appear in the ISI data sets.

5.5.5. INDIVIDUAL RESEARCH FIELD AFFILIATION

A compilation of researcher field affiliations was undertaken to provide an indication of clustering within the citing identity types. A summary is provided in Table 5-4. The boundary specialty of Consciousness was found to have instances of tight and broad clustering of researcher field among citing identity types, which was relatively independent of 'hard' and 'soft' science. For example, MATHEMATICS was found to be tightly clustered, while PHYSICS was widely spread. Similarly NEUROBIOLOGY was tightly clustered, while NEUROPHYSIOLOGY was widely spread. PHILOSOPHY, PSYCHOLOGY and NEUROPSYCHOLOGY were the most widely spread research fields.

Table 5-4 Research field affiliation of authors as obtained from biographical accounts

	Citing identity types						
	1	2	3	4	5	6	7
Anesthesiology	1			1			
Anthropology			1				
Behavioural neuroscience			1		2		
Biophysics					4	1	
Cognitive neuroscience						1	
Cognitive psychology					2		
Cognitive science				1			
Experimental psychologist				1	2	3	1
Linguistics					1		
Mathematics				1	3		
Neurobiology					5		
Neurology				1			
Neurophysiology		1		2	3	2	1
Neuropsychiatry				2			
Neuropsychology			1	2	3		
Pharmacology					1		
Philosophy	1	1	4	4	4	1	
Physics		1		3	2	1	
Physiology			1		1	1	
Psychiatry						1	
Psychobiology					2	1	
Psychology		1	1	1	5	4	
Psychophysiology				1	1	1	
Social psychology				2			

As can be seen in Table 5-4 citing identity types do not have a significant relationship to the field from which the author is primarily affiliated.

5.6. DISCUSSION

5.6.1. APPLICATION OF METHODOLOGY

The citing identity type analysis undertaken has tended to be successful in highlighting the heterogeneous citing behaviours of participants. The high level of self-citing of TART was discovered through analysis of citees recited relative to others, as one example. The extremities of the citing identity types (types 1 and 7) also highlighted researchers with citing attributes that were not shared by the majority of participants. NATSOULAS is a member of citing type 7. He does not co-author, nor does he have a particularly high productivity. He publishes almost exclusively in one journal, of which he is the editor. His citing distribution shows an exceptionally high level of citing a core of authors, of which self-citing is also prominent. These features contribute to a very high Gini value.

An analysis of 97 authors will result in some authors being on the extremities of a distribution, chart or map. Those individuals will be prominent, and in general reasons for their placement at the extremities will be apparent within their research personas. However the ability of citation analysis to inform research personas will be judged successful not on the identification of heterogeneous citing behaviours, but on the ability to discern the behaviours of the greater number of participants. Additionally, the identification of outliers is not feasible using a single citation analysis technique; generally several analyses in combination will be used to confirm the outlying researcher, and large sample sizes are likely to be required to identify them. This process is therefore haphazard and labour intensive, which means that there may be no advantage over existing qualitative research methodologies.

The current chapter addressed both individual-level and field-level methodologies in the analysis of Consciousness researchers. Of the field-level methodologies, the multi-dimensional scaling methodology was found to provide limited information that could not be obtained using an alternative methodology. The results required a high-level understanding of complementary qualitative data to allow reasonable interpretation. Network analysis appears to more efficiently add information to the research persona, without the limitations discovered using multi-dimensional scaling. Analysis focussed on the individual was found to give some information to allow relative comparison of individual researchers. More specific findings were hampered by the nature of the Consciousness boundary specialty, which has a complex composition and lack of stratified sampling that made the evaluation of the newly-generated citing identity types difficult.

5.6.2. RESEARCH PERSONA IDENTIFICATION

Of the drivers for placement within a field, productivity was found to be the strongest. Productivity was moderately related to Gini value, and was identified as driving the shape of the journal co-publication multi-dimensional scaling map. Additionally citing types are strongly correlated with the proportion of unicitations. This is an expected result, as the Gini (a measure of dispersion) will be low for an author that has a large proportion of unicitations. It may have been expected that these specific aspects of the citing identity would be research field-driven, and therefore would further translate into a high correlation between citing type and research field. This was not found overall, only for specific research fields. This may be due to the heterogeneous nature of the data sample and difficulties in defining field; for example, one early career author in the field of neuroscience may have a small number of publications, and a later career researcher a large number of publications. The early career researcher from neuroscience may then be indistinguishable (by citing identity type) from a late-career philosopher. In summary, an open question remains on identifying field through analysis of citing identity characteristics.

The acceptance of the use of citing identities can also be judged on their correlation with existing bibliometric techniques. The level of correlation between field-level and individual-level citation analyses was poor overall for the Consciousness researchers. The usage of information for each of these analyses is substantially different, although the original information is identical; specifically the individual-level citation analysis and generated citing identity types were based on an aggregated measure. As such, there is limited outcome that can be expected from direct analysis of this aggregated number, compared to the richness of analysing citation or citing matrices and other methods derivable from citing identity approaches.

The understanding that is likely to come from analysis of the citing identity is observing patterns and behaviours associated with that aggregated number, i.e. behaviour of “citing identity type”. Therefore the approach of citing identity is undertaken to allow potential generalisation of behaviour across individuals.

The use of the current citing identity type analysis to place an author within their knowledge production units, i.e. trying to determine working relationships between an author and his or her fellow researchers, is likely to be limited, due to the aggregation process required for generating the Gini value. However there is excellent potential for

analysis of individual authors over time, and determining their development in relation to their broader scientific groupings.

5.7. CONCLUSION

The analysis of the boundary specialty of Consciousness was undertaken using the concept of research personas. The success of field-level and individual-level analyses in this study was variable and little correlation was found between them. The most prominent feature of the analysis was the identification of heterogeneous behaviour, which could be correlated with idiosyncrasies in individual's research personas. However, citation analysis will become more useful in the identification of the individual when differentiation is possible between the majority of studied subjects. If this can be achieved, a basis can be established upon which to develop understanding of an individual's communication work practices, through the generation of a research persona. The generation of citing identity type, while not validated in the current study for use in comparative studies, has potential for utility in this endeavour with further methodological development

6. CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

6.1. CONCLUDING REMARKS

The preliminary questions in this thesis relate to the use of citations to assist in understanding the role of the situated individual in knowledge production. More specifically, the use of types and identities derived from citation analysis, and whether these can be adequately described by citation theories, need to be understood. The use of Consciousness to perform these studies was on the basis of establishing a participant set of individuals having diverse research field affiliations, and an establishment of commonality of research interest that posed an interesting subject for the analysis of research persona that was concurrently undertaken.

The differentiation of authors according to their citing identities and types was successful. Following the concepts proposed and developed by White, a relationship was determined between Gini and Citee-citation ratio. The relation was tested and provisionally determined to be non-trivial, and furthermore current bibliometric “laws” applied to the relation yielded no direct explanation of the phenomenon. The result was the establishment of a baseline and methodology on which to base the definition of citing identity type. Determination of concentration of citing profiles through calculation of Gini was a successful component of this research, and may prove to be a significant method for the understanding of authors through citation, although as an aggregated measure it must be accepted there is loss of detail. The broad applicability of the discovered relation is as yet undetermined; however the diverse nature of Consciousness gives confidence that extension of this research into new fields and specialties may be successful. Data-generated citing identity types were established on the basis of this relation, providing an “independent” establishment of type.

Further analysis of the relation amongst Consciousness authors was undertaken, to determine whether citation analysis, and the newly generated citing identity types, could be used to understand the research practices of individuals. Consciousness provided a suitable field for the establishment of a citing identity type, but proved to be too diverse for subsequent verification. The strongest underlying factor found to correlate to citing identity type was productivity, but further influences were difficult to determine. Analysis of research personas was undertaken using citing identity type and a range of other citation, biographic and bibliographic information, using already-developed bibliometric methods. This provided both field-level and individual-level

analysis. In the Consciousness data set no correlation could be found between analyses at these levels. At the conclusion of this research, a link between research practices of individuals and citing identities must be described as weak, partially due to the complexity of the object of study.

The concept of citation theory being applied to characterization of the individual knowledge producer arose from being central to the communication processes in academic careers. To determine what citation theories will require to adequately address aspects of research persona, further methodological development and verification of citation theories will be required. The utility of citing type may grow if standard types can be identified, and their applicability is confirmed in a variety of research environments. If successful, this will provide added richness to the understanding of the individual researcher and their scientific work practices.

6.2. RECOMMENDATIONS FOR FURTHER RESEARCH

The current work has utilized a boundary specialty, to determine broad applicability of the use of standardized citing identity types in analysing researchers. Areas of further work are recommended, both in methodology development in citation analysis, and in the development of research personas as a means to access identity.

The limitation of Consciousness, once the relation between Gini and Citee-citation ratio had been discovered, was the complex mix of researcher characterization including the overlapping of fields, researcher career lengths, and publication levels. The verification of the relation as an applicable basis for the generation of citing identity types is a clear requirement for following research. Success in this endeavour is dependent on identifying “simple” fields or research specialties. Principally, it will be important to determine the characteristics of a set of similar researchers in field, career length, publication and standing. In principle these could form a single citing identity type, and therefore may not form a distinct relation between Gini and Citee-citation ratio. Subsequently, the influence of time on the citing profile can be determined. This may be either environmental (changing of citing practices over time) or personal (development of citing behaviour over the length of career). Finally, confirmation of the trend evident in the current study of stratification of researchers according to productivity requires further investigation.

Further work on scholarly identities and research personas will be possible with the further development of methodology outlined above. This will allow the use of combinations of qualitative and quantitative techniques in establishing a richer

understanding of scholarly identity. An example of this is the examination of the integration and specialisation of researchers use of 'common bodies of knowledge' within their representational work practices.

BIBLIOGRAPHY

- Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 50, 179-211.
- Amsterdamska, O., & Leydesdorff, L. (1989). Citations - Indicators of Significance. *Scientometrics*, 15(5-6), 449-471.
- Anderberg, M. R. (1973). *Cluster analysis for applications*. New York: Academic Press.
- Anderson, T. (2004, 1st - 3rd September). *Relevance as process: judgments in the context of scholarly research*. Paper presented at the Information Seeking in Context (ISIC) 2004, Dublin, Ireland.
- Anderson, T. (2005). Relevance as process: judgments in the context of scholarly research. *Information Research*, 10(2), paper 226 [Available at <http://InformationR.net/ir/210-222/paper226.html>].
- Argyrous, G. (2000). *Statistics for social and health research: with a guide to SPSS*. London: Sage Publications.
- ASSC: Association for the Scientific Studies of Consciousness (No date, No date). [Online]. California Institute of Technology. Retrieved 11 June, 2004, from the World Wide Web: <http://assc.caltech.edu/index.htm>
- Baldi, S. (1998). Normative versus social constructivist processes in the allocation of citations: a network analytic model. *American Sociological Review*, 63(6), 829-846.
- Barber, B. (1952). *Science and the social order*. New York: The Free Press.
- Bavelas, J. B. (1978). The social psychology of citations. *Canadian Psychological Review*, 19(3), 158-163.
- Bazerman, C. (1983). Scientific writing as a social act: a review of the literature of the sociology of science. In P. V. Anderson & R. J. Brockman & C. R. Miller (Eds.), *New Essays in Technical and Scientific Communication* (pp. 156-184). Farmingdale, N.Y: Baywood.
- Becher, T. (1989). *Academic Tribes and Territories: intellectual enquiry and the cultures of disciplines*. Milton Keynes [England] ; Bristol, PA., USA: The Society for Research into Higher Education & Open University Press.
- Berkenkotter, C., & Huckin, T. (1995). *Genre knowledge in disciplinary communication*. Hillsdale, NJ: Lawrence Erlbaum.
- Borgman, C. L., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36, 3-72.
- Braam, R. R., Moed, H. F., & Van Raan, A. F. J. (1991). Mapping of science by combined co-citation and word analysis. 1. Structural aspects. *Journal of the American Society for Information Science*, 42(4), 233-251.
- Brooks, T. A. (1985). Private acts and public objects: an investigation of citer motivations. *Journal of the American Society for Information Science*, 36(4), 223-229.
- Brooks, T. A. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science*, 37, 34-36.
- Budd, J. M. (1999). Citations and knowledge claims: sociology of knowledge as a case in point. *Journal of Information Science*, 25(4), 265-274.
-

-
- Burrell, Q. L. (1991). The Bradford distribution and the Gini index. *Scientometrics*, 21(2), 181-194.
- Callon, M. (1980). Struggles and negotiations to define what is problematic and what is not: the socio-logic of translation. In K. D. Knorr & R. Krohn & R. Whitley (Eds.), *The social process of scientific investigation* (pp. 197-219). Dordrecht, Holland: D. Reidel Pub. Co.
- Callon, M. (1986). Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. In J. Law (Ed.), *Power, action and belief: a new sociology of knowledge?* (pp. 196-223). London: Routledge.
- Callon, M., Law, J., & Rip, A. (Eds.). (1986). *Mapping the dynamics of science and technology*. London: Macmillan.
- Case, D. O., & Higgins, G. M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science*, 51(7), 635-645.
- Chubin, D. (1981). Constructing and reconstructing scientific reality: a meta-analysis. *Newsletter of the International Society for the Sociology of Knowledge*, 7, 22-28.
- Chubin, D., & Connolly, T. (1982). Research Trails and Science Policies: local and extra-local negotiation of scientific work. In N. Elias & H. Martins & R. Whitley (Eds.), *Scientific establishments and hierarchies* (Vol. 6, pp. 293-377). Dordrecht, Holland: Reidel.
- Chubin, D. E. (1976). The Conceptualization of Scientific Specialties. *Sociological Quarterly*, 17(4), 448-476.
- Chubin, D. E., & Moitra, S. D. (1975). Content analysis of references: adjunct or alternative to citation counting? *Social Studies of Science*, 5(4), 423-441.
- Cole, J. R. (1970). Patterns of Intellectual Influence in Scientific Research. *Sociology of Education*, 43(4), 377-403.
- Cole, S. (1975). The growth of scientific knowledge: theories of deviance as a case study. In L. Coser (Ed.), *The idea of social structure* (pp. 175-220). New York: Harcourt Brace Jovanovich.
- Courtial, J. P. (2002). Analysis of social representations in action based on words associated by scientific articles. *European Review of Applied Psychology*, 52(3-4), 221-230.
- Cozzens, S. E. (1981). Taking the measure of science: a review of citation theories. *Newsletter of the International Society for the Sociology of Knowledge*, 7, 16.
- Cozzens, S. E. (1985). Comparing the sciences: citation context analysis of papers from neuropharmacology and the sociology of science. *Social Studies of Science*, 15(1), 127-153.
- Cozzens, S. E. (1989). What do citations count? The rhetoric-first model. *Scientometrics*, 15(5-6), 437-447.
- Crick, F. (1993). *The astonishing hypothesis: the scientific search for the soul*. New York: Charles Scribner's Sons.
- Cromby, J. (2004). Between Constructionism and Neuroscience: The Societal Co-constitution of Embodied Subjectivity. *Theory & Psychology*, 14(6), 797-821.
- Cronin, B. (1981). The Need for a Theory of Citing. *Journal of Documentation*, 37(1), 16-24.
-

-
-
- Cronin, B. (1982). Norms and functions in citation: the view of journal editors and referees in psychology. *Social Science Information Studies*, 2(2), 65-78.
- Cronin, B. (1984). *The citation process : the role and significance of citations in scientific communication*. London: Taylor Graham.
- Cronin, B. (2000). Semiotics and evaluative bibliometrics. *Journal of Documentation*, 56(4), 440-453.
- Cronin, B. (2004). Normative shaping of scientific practice: The magic of Merton. *Scientometrics*, 60(1), 41-46.
- Cronin, B., & Shaw, D. (2002). Identity-creators and image-makers: using citation analysis and thick description to put authors in their place. *Scientometrics*, 54(1), 31-49.
- de Quincey, C. (2006). Switched-on Consciousness: Clarifying What It Means. *Journal of Consciousness Studies*, 13(4), 7-12.
- Dennett, D. (1993). *Consciousness Explained*. London: Penguin.
- Editorial: the future of consciousness studies. (1997). *Journal of Consciousness Studies*, 4(5-6), 385-388.
- Egghe, L., & Rousseau, R. (1990). *Introduction to Informetrics: quantitative methods in library, documentation and information science*. Amsterdam: Elsevier.
- Egghe, L., & Rousseau, R. (2003). Personal correspondence.
- Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). London: Arnold.
- Galbraith, D. (1999). Writing as a knowledge-constituting process. In M. Torrance & D. Galbraith (Eds.), *Knowing what to write: Conceptual processes in text production* (Vol. 4, pp. 139-159). Amsterdam: Amsterdam University Press.
- Garfield, E. (1977). *Can citation indexing be automated?* [PDF]. ISI Press. Retrieved, from the World Wide Web:
<http://www.garfield.library.upenn.edu/essays/V1p084y1962-73.pdf>
- Gergen, K. J. (1982). *Toward transformation in social knowledge*. New York: Springer-Verlag.
- Gergen, K. J. (1991). *The saturated self: dilemmas of identity in contemporary life*. New York: Basic Books.
- Gieryn, T. F. (1978). Problem retention and problem change in science. *Sociological Inquiry*, 48(3-4), 96-115.
- Gieryn, T. F., & Hirsh, R. F. (1983). Marginality and Innovation in Science. *Social Studies of Science*, 13(1), 87-106.
- Gilbert, N. (1976). The transformation of research findings into scientific knowledge. *Social Studies of Science*, 6(3&4), 281-306.
- Gilbert, N. (1977). Referencing as persuasion. *Social Studies of Science*, 7(1), 113-122.
- Gilbert, N., & Woolgar, S. (1974). The quantitative study of science: an examination of the literature. *Science Studies*, 4, 279-294.
- Gläser, J. (2001a). Macrostructures, careers and knowledge production: a neoinstitutionalist approach. *International Journal of Technology Management*, 22(7-8), 698-715.
- Gläser, J. (2001b). *Scientific specialties as the (currently missing) link between scientometrics and the sociology of science*. Paper presented at the 8th
-
-

International Conference on Scientometrics & Informetrics ISSI-2001, Sydney, Australia.

- Hameroff, S. (1994, 1996). *Toward a Science of Consciousness: The First Tucson Discussions and Debates*. Paper presented at the Toward a Scientific Basis for Consciousness, "Tucson I", April 12-17, 1994, Tucson, Arizona.
- Hargens, L. L. (2000). Using the literature: reference networks, reference contexts, and the social structure of scholarship. *American Sociological Review*, 65(6), 846-865.
- Harsanyi, M. A. (1993). Multiple authors, multiple problems - bibliometrics and the study of scholarly collaboration: a literature review. *Library & Information Science Research*, 15, 325-354.
- Harter, S. P. (1992). Psychological relevance and information space. *Journal of the American Society for Information Science*, 43(9), 602-615.
- Harter, S. P., Nisonger, T. E., & Weng, A. (1993). Semantic relationships between cited and citing journals in library and information science journals. *Journal of the American Society for Information Science*, 44(9), 543-552.
- Herdan, G. (1960). *Type-token mathematics: a textbook of mathematical linguistics*. 'S-Gravenhage: Mouton & Co.
- Hicks, D., & Potter, J. (1991). Sociology of scientific knowledge: a reflexive citation analysis of science disciplines and disciplining science. *Social Studies of Science*, 21(3), 459-501.
- Howard, D. L. (1994). Pertinence as Reflected in Personal Constructs. *Journal of the American Society for Information Science*, 45(3), 172-185.
- Hyland, K. (1999). Academic attribution: citation and the construction of disciplinary knowledge. *Applied Linguistics*, 20(3), 341-367.
- Journal of Consciousness Studies* (no date, no date). Imprint Academic. Retrieved 29 August, 2006, from the World Wide Web: <http://www.imprint.co.uk/jcs.html>
- Kaplan, N. (1965). The norms of citation behaviour: prolegomena to the footnote. *American Documentation*, 16, 179-184.
- Kaszniak, A. (No date). *Al Kaszniak*. University of Arizona. Retrieved 26 July, 2005, from the World Wide Web: <http://web.arizona.edu/~cnl/alkaszniak.htm>
- Kinnear, P. R., & Gray, C. D. (1999). *SPSS for Windows made simple* (3rd ed.). East Sussex: Psychology Press.
- Kreuzman, H. (2001). A co-citation analysis of representative authors in philosophy: Examining the relationship between epistemologists and philosophers of science. *Scientometrics*, 51(3), 525-539.
- Kuhn, T. S. (1963). The essential tension: tradition and innovation in scientific research. In C. W. Taylor & F. Barron (Eds.), *Scientific creativity: its recognition and development* (pp. 341-354). New York: Wiley.
- Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Murray, S. S., Martimbeau, N., & Elwell, B. (2005). The bibliometric properties of article readership information. *Journal of the American Society for Information Science and Technology*, 56(2), 111-128.
- Langham, T. (1995). Consistency in Referencing. *Journal of Documentation*, 51(4), 360-369.
- Latour, B. (1987). *Science in action: how to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.
-

-
-
- Latour, B., & Woolgar, S. (1979). *Laboratory Life: the social construction of scientific facts*. Beverly Hills, CA: Sage.
- Laudel, G. (2001). Collaboration, creativity and rewards: why and how scientists collaborate. *International Journal of Technology Management*, 22(7-8), 762-781.
- Laudel, G. (2003). Studying the brain drain: Can bibliometric methods help? *Scientometrics*, 57(2), 215-237.
- Law, J., & Williams, R. J. (1982). Putting Facts Together: A Study of Scientific Persuasion. *Social Studies of Science*, 12(4 (Theme Section: Laboratory Studies)), 535-558.
- Leimkuhler, F. F. (1970). The Bradford distribution. In T. Saracevic (Ed.), *Introduction to information science* (pp. 509-514). New York: Bowker.
- Lenoir, T. (1997a). *Instituting science: the cultural production of scientific disciplines*. Stanford: Stanford University Press.
- Lenoir, T. (1997b). Introduction, *Instituting Science: The Cultural Production of Scientific Disciplines* (pp. 1-21). Stanford: Stanford University Press.
- Lenoir, T. (1998). Inscription Practices and Materialities of Communication, *Inscribing Science: Scientific Texts and the Materialities of Communication* (pp. 1-19). Stanford: Stanford University Press.
- Lewis-Beck, M. S. (1980). *Applied regression: an introduction* (Vol. 07-022). Newbury Park: Sage Publications.
- Leydesdorff, L. (1987). Towards a Theory of Citation. *Scientometrics*, 12(5-6), 305-309.
- Leydesdorff, L. (1989). The Relations between Qualitative Theory and Scientometric Methods in Science and Technology Studies - Introduction to the Topical Issue. *Scientometrics*, 15(5-6), 333-347.
- Leydesdorff, L. (1993). The impact of citation behaviour on citation structure. In A. Van Raan & R. E. de Bruin & H. F. Moed & A. J. Nederhof & R. Tijssen (Eds.), *Science and Technology in a Policy Context* (pp. 289-300). Leiden: DSWO Press, Leiden University.
- Leydesdorff, L. (1995). *The challenge of scientometrics: the development, measurement and self-organisation of scientific communications* (2nd ed 2001 ed.). Leiden: DSWO Press, Leiden University.
- Leydesdorff, L. (1998). Theories of citation? *Scientometrics*, 43(1), 5-25.
- Leydesdorff, L. (2005). Anticipatory systems and the processing of meaning: a simulation study inspired by Luhmann's Theory of Social Systems. *Journal of Artificial Societies and Social Simulation*, 8(2), <http://jasss.soc.surrey.ac.uk/8/2/7.html>.
- Leydesdorff, L. (2006). *Scientific Communication and Cognitive Codification: Social Systems Theory and the Sociology of Scientific Knowledge* [PDF]. Amsterdam School of Communications Research (ASCoR), University of Amsterdam. Retrieved 2 February, 2006, from the World Wide Web: <http://users.fmg.uva.nl/lleydesdorff/ejts06/index.htm> - [Forthcoming in 2007 in the *European Journal of Social Theory* 10 (3)]
- Leydesdorff, L., & Amsterdamska, O. (1990). Dimensions of citation analysis. *Science, Technology and Human Values*, 15, 205-335.
-
-

-
- Leydesdorff, L., & van den Besselaar, P. (1997). Scientometrics and communication theory: towards theoretically informed indicators. *Scientometrics*, 38(1), 155-174.
- Leydesdorff, L., & Vaughan, L. (2005 preprint). *Co-occurrence matrices and their applications in information science: extending ACA to the web environment* [PDF]. Retrieved 19 July, 2005, from the World Wide Web: <http://users.fmg.uva.nl/lleydesdorff/aca/aca.pdf>
- Leydesdorff, L., & Wouters, P. (1999). Between texts and contexts: advances in theories of citation? (A rejoinder). *Scientometrics*, 44, 169-182.
- Lievrouw, L. A. (1989). The invisible college reconsidered: bibliometrics and the development of scientific communication theory. *Communication Research*, 16(5), 615-628.
- Lievrouw, L. A. (1990). Reconciling structure and process in the study of scholarly communication. In C. L. Borgman (Ed.), *Scholarly communication and bibliometrics* (pp. 59-69). Newbury Park, CA: Sage.
- Lievrouw, L. A. (1992). Communication, representation, and scientific knowledge: A conceptual framework and case study. *Knowledge and Policy: The International Journal of Knowledge Transfer and Utilization*, 5(1), 6-28.
- Lievrouw, L. A. (1996). Constructing research narratives and establishing scholarly identities: Properties and propositions. In H. B. Mokros (Ed.), *Interaction and Identity* (Vol. 5, pp. 215-235). New Brunswick, NJ: Transaction.
- Liu, M. X. (1993). Progress in Documentation - the Complexities of Citation Practice - a Review of Citation Studies. *Journal of Documentation*, 49(4), 370-408.
- Losee, R. M. (1996). Text windows and phrases differing by discipline, location in document, and syntactic structure. *Information Processing & Management*, 32(6), 747-767.
- Luukkonen, T. (1997). Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis. *Scientometrics*, 38(1), 27-37.
- MacRoberts, M. H., & MacRoberts, B. R. (1989). Problems of citation analysis: a critical review. *Journal of the American Society for Information Science*, 40(5), 342-349.
- McCain, K. W. (1990). Mapping authors in intellectual space: a technical overview. *Journal of the American Society for Information Science*, 41(6), 433-443.
- McCain, K. W. (1998). Neural networks research in context: A longitudinal journal cogitation analysis of an emerging interdisciplinary field. *Scientometrics*, 41(3), 389-410.
- Merton, R. K. (1959). *Sociology today: problems and prospects*. New York: Basic Books.
- Merton, R. K. (1973). *The sociology of science: theoretical and empirical investigations*. Chicago: University of Chicago Press.
- Merton, R. K. (1988). The Matthew Effect in Science, II: Cumulative advantage and the symbolism of intellectual property. *ISIS*, 79, 606-623.
- Merton, R. K. ([1942] 1973). The Normative Structure of Science. In R. K. Merton & N. W. Storer (Eds.), *The Sociology of Science: theoretical and empirical investigations*. Chicago: The University of Chicago Press.
-

-
- Merton, R. K. ([1957] 1973). Priorities in scientific discovery. In N. W. Storer (Ed.), *The Sociology of Science: theoretical and empirical investigations* (pp. 287-324). Chicago: The University of Chicago Press.
- Merton, R. K. ([1968] 1973). Behaviour patterns of scientists. In N. W. Storer (Ed.), *The Sociology of Science: theoretical and empirical investigations* (pp. 325-342). Chicago: The University of Chicago Press.
- Mind-body problem* (2005, 18 June 2005). Wikimedia Foundation. Retrieved 20 July, 2005, from the World Wide Web: http://en.wikipedia.org/wiki/Mind-body_problem
- Mitroff, I. I. (1974). Norms and Counter-Norms in a Select Group of the Apollo Moon Scientists: A Case Study of the Ambivalence of Scientists. *American Sociological Review*, 39(4), 579-595.
- Moed, H. F., & Garfield, E. (2004). In basic science the percentage of 'authoritative' references decreases as bibliographies become shorter. *Scientometrics*, 60(3), 295-303.
- Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science*, 5(1), 86-92.
- Mulkay, M. J. (1976a). Methodology in the sociology of science: some reflections on the study of Radio Astronomy. In G. Lemaine & R. MacLeod & M. J. Mulkay & P. Weingart (Eds.), *Perspectives in the emergence of scientific disciplines* (pp. 208-220). Mouton: Paris.
- Mulkay, M. J. (1976b). Norms and Ideology in Science. *Social Science Information*, 15(4-5), 637-656.
- Mulkay, M. J., Gilbert, G. N., & Woolgar, S. (1975). Problem areas and research networks in science. *Sociology*, 9, 187-203.
- Myers, G. (1993). The social construction of two biologists' articles. In E. Messer-Davidow & D. Shumway & D. Sylvan (Eds.), *Knowledges: Historical and Critical Studies in Disciplinarity* (pp. 327-367). Charlottesville: University Press of Virginia.
- Nicolaisen, J. (2003). *The social act of citing: towards new horizons in citation theory*. Paper presented at the Proceedings of the 66th American Society for Information Science and Technology Annual Meeting, October 19-22 2003, Long Beach, California.
- Nicolaisen, J. (2004). *Social behaviour and scientific practice - missing pieces of the citation puzzle*. Unpublished PhD, Royal School of Library and Information Science, Denmark.
- Noyons, E. C. M. (2001). Bibliometric mapping of science in a science policy context. *Scientometrics*, 50(1), 83-98.
- Noyons, E. C. M., Moed, H. F., & Luwel, M. (1999). Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. *Journal of the American Society for Information Science*, 50(2), 115-131.
- Paul, D. (2000). In citing chaos: a study of the rhetorical use of citations. *Journal of Business and Technical Communication*, 14(2), 185-222.
- Peters, H. P. F., Braam, R. R., & Van Raan, A. F. J. (1995). Cognitive resemblance and citation relations in chemical- engineering publications. *Journal of the American Society for Information Science*, 46(1), 9-21.
- Petranker, J. (2001). Who Will Be the Scientists? A Review of B. Alan Wallace's 'The Taboo of Subjectivity'. *Journal of Consciousness Studies*, 8(11), 83-90.
-

-
- Polanyi, M. (1958). *Personal knowledge*. London: Routledge and Kegan Paul.
- Rousseau, R. (2000, 23-25 October). *Concentration and evenness measures as macro-level scientometric indicators*. Paper presented at the Second International Symposium on Quantitative Evaluation of Research Performances and Sixth National Conference on Scientometrics and Informetrics, Shanghai, China.
- Rousseau, R. (2001). Concentration and evenness measures as macro-level scientometric indicators. (English Translation). In J. Guo-Hua (Ed.), *Research and university evaluation* (pp. 72-88): Red Flag Publishing House.
- Scaruffi, P. (no date). *Quantum Consciousness* [Online]. Retrieved 20 July, 2005, from the World Wide Web: <http://www.thymos.com/science/qc.html>
- Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: toward a dynamic, situational definition. *Information Processing & Management*, 26, 755-776.
- Schmitt, F. (Ed.). (1994). *Socializing epistemology: the social dimensions of knowledge*. Langam. MD.: Rowman & Littlefield.
- Schwechheimer, H., & Winterhager, M. (2001). Mapping interdisciplinary research fronts in neuroscience: a bibliometric view to retrograde amnesia. *Scientometrics*, 51(1), 311-318.
- Shadish, W. R., Tolliver, D., Gray, M., & Gupta, S. K. (1995). Author judgments about works they cite: three studies from psychology journals. *Social Studies of Science*, 25(3), 477-498.
- Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24, 265-269.
- Small, H. (1978). Cited documents as concept symbols. *Social Studies of Science*, 8, 327-340.
- Small, H. (1980). Co-citation analysis and the structure of paradigms. *Journal of Documentation*, 36, 183-196.
- Small, H. (1986). The synthesis of specialty narratives from co-citation clusters. *Journal of the American Society for Information Science*, 37(3), 97-110.
- Small, H. (1987). The significance of bibliographic references. *Scientometrics*, 12, 339-341.
- Small, H. (1998). Citations and consilience in science - Comments on theories of citation? *Scientometrics*, 43(1), 143-148.
- Small, H. (1999a). A passage through science: Crossing disciplinary boundaries. *Library Trends*, 48(1), 72-108.
- Small, H. (1999b). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50(9), 799-813.
- Small, H. (2003). Paradigms, citations and maps of science: a personal history. *Journal of the American Society for Information Science and Technology*, 54(5), 394-399.
- Small, H. (2004a). On the shoulders of Robert Merton: Towards a normative theory of citation. *Scientometrics*, 60(1), 71-79.
- Small, H. (2004b). Why authors think their papers are highly cited. *Scientometrics*, 60(3), 305-316.
-

-
-
- Small, H., & Griffith, B. (1974). The structure of scientific literatures 1: identifying and graphing specialties. *Science Studies*, 4(1), 17-40.
- Star, S. L. (1983). Simplification in scientific work: an example from neuroscience research. *Social studies of Science*, 13(2), 205-228.
- Star, S. L. (1995). Politics of formal representations: wizards, gurus, and organizational complexity. In S. L. Star (Ed.), *Ecologies of knowledge: work and politics in science and technology* (pp. 88-118). Albany, N.Y.: State University of New York Press.
- Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19, 387-420.
- Stegmann, J., & Grohmann, G. (2001). Citation rates, knowledge export and international visibility of dermatology journals listed and not listed in the Journal Citation Reports. *Scientometrics*, 50(3), 483-502.
- Stets, J. E., & Burke, P. J. (2005). A Sociological Approach to Self and Identity. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 128-152). New York: Guilford.
- Stewart, J. A. (1983). Achievement and ascriptive processes in the recognition of scientific articles. *Social Forces*, 62, 166-189.
- Storer, N. W. (1966). *The social system of science*. New York: Holt, Rinehart and Winston.
- Sutherland, K. (1998). TUCSON III — A Personal View. *Journal of Consciousness Studies*, 5(4).
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York, NY: Harper Collins College publishers.
- Toward a Science of Consciousness*. (2000). Paper presented at the Toward a Science of Consciousness, "Tucson 2000", April 10-15, 2000, Tucson, Arizona.
- Toward a Science of Consciousness: The First Tucson Discussions and Debates*. (1994, 1996). Paper presented at the Toward a Scientific Basis for Consciousness, "Tucson I", April 12-17, 1994, Tucson, Arizona.
- Toward a Science of Consciousness: The Second Tucson Discussions and Debates*. (1996, 1998). Paper presented at the Toward a Science of Consciousness, "Tucson II", April 8-13, 1996, Tucson, Arizona.
- Toward a Science of Consciousness: The Third Tucson Discussions and Debates*. (1998, 1999). Paper presented at the Toward a Science of Consciousness, "Tucson III", April 27-May 2, 1998, Tucson, Arizona.
- van den Besselaar, P., & Leydesdorff, L. (1996). Mapping change in scientific specialities: a scientometric reconstruction of the development of Artificial Intelligence. *Journal of the American Society of Information Science*, 47(6), 415-436.
- Van der Veer Martens, B. (2001). Do citation systems represent theories of truth? *Information Research*, 6(2), [Available at: <http://InformationR.net/ir/6-2/paper92.html>].
- Van der Veer Martens, B., & Goodrum, A. A. (2006). The diffusion of theories: A functional approach. *Journal of the American Society for Information Science and Technology*, 57(3), 330-341.
-
-

-
- van Raan, A. F. J. (1998). In matters of quantitative studies of science the fault of theorists is offering too little and asking too much - Comments on theories of citation? *Scientometrics*, 43(1), 129-139.
- Vinkler, P. (1993). Research Contribution , Authorship and Team Cooperativeness. *Scientometrics*, 26(1), 213-230.
- Wang, P., & Soergel, D. (1998). A cognitive model of document use during a research project. Study 1: document selection. *Journal of the American Society for Information Science*, 49, 115-133.
- Wang, P., & White, M. D. (1999). A cognitive model of document use during a research project. Study II. Decisions at the reading and citing stages. *Journal of the American Society for Information Science and Technology*, 50(2), 98-114.
- White, H. D. (2000). Toward ego-centered citation analysis, *Web of Knowledge - a Festschrift in Honor of Eugene Garfield* (pp. 475-496).
- White, H. D. (2001a). Author-centered bibliometrics through CAMEOs: Characterizations automatically made and edited online. *Scientometrics*, 51(3), 607-637.
- White, H. D. (2001b). Authors as citers over time. *Journal of the American Society for Information Science and Technology*, 52(2), 87-108.
- White, H. D. (2004a). Citation analysis and discourse analysis revisited. *Applied Linguistics*, 25(1), 89-116.
- White, H. D. (2004b). Reward, persuasion, and the Sokal Hoax: A study in citation identities. *Scientometrics*, 60(1), 93-120.
- White, H. D., & McCain, K. W. (1997). Visualization of literatures. *Annual Review of Information Science and Technology*, 32, 99-168.
- 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 for Information Science*, 49(4), 327-355.
- White, H. D., Wellman, B., & Nazer, N. (2004). Does citation reflect social structure? Longitudinal evidence from the "Globenet" interdisciplinary research group. *Journal of the American Society for Information Science and Technology*, 55(2), 111-126.
- White, M. D., & Wang, P. (1997). A qualitative study of citing behavior: contributions, criteria, and metalevel documentation concerns. *Library Quarterly*, 67(2), 122-154.
- Whitehead, C. (2004). Everything I believe might be a delusion. Whoa! - Tucson 2004: ten years on, and are we any nearer to a Science of Consciousness? *Journal of Consciousness Studies*, 11(12), 68-88.
- Whitley, R. D. (1969). Communication nets in science: status and citation patterns in animal physiology. *Sociological Review*, 17, 219-233.
- Whitley, R. D. (1980). The context of scientific investigation. In K. D. Knorr & R. Krohn & R. Whitley (Eds.), *The social process of scientific investigation* (pp. 297-321). Dordrecht, Holland: D. Reidel Pub. Co.
- Whitley, R. D. (1984). *The intellectual and social organisation of the sciences*. Oxford: Clarendon Press.
- Wilson, C. S. (1999). Informetrics. *Annual Review of Information Science and Technology*, 34, 107-247.
-

-
- Woolgar, S. (1976). Writing an intellectual history of scientific development: the use of discovery accounts. *Social Studies of Science*, 6, 395-422.
- Wouters, P. (1997). *The signs of science*. Paper presented at the Sixth International Conference of the International Society for Scientometrics and Informetrics, Jerusalem, Israel.
- Wouters, P. (1999). *The citation culture*. Unpublished Ph.D Thesis, University of Amsterdam, Amsterdam.
- Zuckerman, H. A. (1968). Patterns of name-ordering among scientific papers: a study of social symbolism and its ambiguity. *American Journal of Sociology*, 74, 276.
- Zuckerman, H. A. (1978). Theory choice and problem choice in science. *Sociological Inquiry*, 48(3-4), 65-95.
- Zuckerman, H. A. (1984). Norms and Deviant-Behavior in Science. *Science Technology & Human Values*, 46, 7-13.
- Zuckerman, H. A. (1987). Citation analysis and the complex problem of intellectual influence. *Scientometrics*, 12(5-6), 329-338.
- Zuckerman, H. A. (1988). The sociology of science. In N. J. Smelser (Ed.), *Handbook of sociology* (pp. 511-574). Newbury Park: Sage Publications.
- Zuckerman, H. A., & Cole, J. R. (1994). Research strategies in science: a preliminary inquiry. *Creativity Research Journal*, 7(3-4), 391-405.
- Zuckerman, H. A., & Merton, R. K. (1971). Patterns of evaluation in science: institutionalization, structure and functions of the referee system. *Minerva*, 9, 66-100.

APPENDIX A. AUTHOR CODES

AUTHOR	AUTHOR CODE	AUTHOR	AUTHOR CODE
AHERN GL (Geoffrey)	AGL	LEVINE DS (Daniel)	LDS
ARMSTRONG DM (David)	ADM	LEVINE J (Joseph)	LEJ
BAARS BJ (Bernard)	BBJ	LIBET B (Benjamin)	LIB
BECK F (Friedrich)	BEF	LLINAS RR (Rodolfo)	LRR
BLOCK NJ (Ned)	BNJ	LOCKWOOD M (Michael)	LMJ
BLOOM P (Paul)	BLP	LOGOTHETIS NK (Nikos)	LNK
BOGEN JE (Joseph)	BJE	MANGAN B (Bruce)	MBB
	BDJ	MARCEL AJ (Anthony/Tony)	MAJ
BOHM D (David)		MARSHALL IN (Ian)	MIN
CHALMERS DJ (David)	CDJ	MILNER AD (David)	MAD
CHURCHLAND PM (Paul)	CPM		MOG
CHURCHLAND PS (Patricia)	CPS	MORUZZI G (Giuseppe)	
CONRAD ME (Michael)	CME	NAGEL T (Thomas)	NAG
COWEY A (Alan)	COA	NATSOULAS T (Thomas)	NAT
CRICK FHC (Francis)	CFH	NEWMAN J (James)	NEJ
DAMASIO AR (Antonio)	DAR	PENROSE R (Roger)	PRJ
DAVIDSON RJ (Richard)	DRJ	PLUM F (Fred)	PLF
DENNETT DC (Daniel)	DDC	POPPER KR (Karl)	PKR
ECCLES JC (John)	EJC	POSNER MI (Michael)	PMI
EDELMAN GM (Gerald)	EGM	PRIBRAM KH (Karl)	PKH
FENIGSTEIN A (Allen)	FEA	REIMAN EM (Eric)	REM
FLANAGAN O (Owen)	FOJ	RIZZOLATTI G (Giacomo)	RIG
FLOHR H (Hans)	FLH	ROSENTHAL DM (David)	ROD
FRANKS NP (Nicholas)	FNP	ROSSETTI Y (Yves)	ROY
FROHLICH H (Herbert)	FRH	RUMBAUGH DM (Duane)	RDM
GALIN D (David)	GAD	SAGI D (Dov)	SAD
GAZZANIGA MS (Michael)	GMS	SATARIC MV	SMV
GOODALE MA (Melvyn)	GMA	SCHACTER DL (Daniel)	SDL
GRAY CM (Charles)	GCM	SCHWARTZ GE (Gary)	SGE
GRAY JA (Jeffery)	GJA	SCOTT AC (Alwyn)	SAC
GREENFIELD SA (Susan)	GSA	SEARLE JR (John)	SJR
HAIER RJ (Richard)	HRJ	SHALLICE T (Tim)	SHT
HAMEROFF SR (Stuart)	HSR	SHEPARD RN (Roger)	SRN
HARTH E (Erich)	HAE	SINGER W (Wolf)	SIW
HEBB DO (Donald)	HDO	SPERRY RW (Roger)	SRW
HOBSON JA (John)	HJA	SQUIRES EJ (Euan)	SEJ
HODGSON DH (David)	HDH	STAPP HP (Henry)	SHP
HUBBARD TL (Timothy)	HTL	STERIADE M (Mircea)	STM
JACKENDOFF R (Ray)	JRS	STOERIG P (Petra)	STP
JIBU M (Mari)	JIM	TART CT (Charles)	TCT
JOHNSONLAIRD PN (Philip)	JPN		TJG
		TAYLOR JG (John)	
KASZNIAK AW (Alfred)	KAW	TULVING E (Endel)	TEE
KIHLSTROM JF (John)	KJF	TUSZYNSKI JA (Jack)	TJA
KOCH C (Christof)	KOC	VARELA FJ (Francisco)	VFJ
KONIG P (Peter)	KOP	VELMANS M (Max)	VML
LABERGE SP (Stephen)	LSP	WALSH R (Roger N)	WRN
LANE RD (Richard)	LRD	WATT RC (Richard)	WRC
	LCD	WEISKRANTZ L (Lawrence/Larry)	WEL
LAUGHLIN CD (Charles)			
LEOPOLD DA (David)	LDA	WOLF FA (Fred)	WFA
LEVINE DS (Daniel)	LDS	ZAIDEL E (Eran)	ZAE

Authors that only appear in the author co-citation analysis:
Wilder Penfield (1891-1976), Sigmund Freud (1856-1939), William James (1842-

APPENDIX B. PROGRAM CODE

APPENDIX B.1. CODE FOR BIBLIOGRAPHIC EXTRACTION PROGRAM⁶¹

```
// Thesis1.cpp : Strips title, journal, co-author, publication year, cited references,
// number of reference information from a text file
// and places them in separate output text files.
//

#include "stdafx.h"
#include <stdio.h>
#include <string.h>
#include <malloc.h>
#include <stdlib.h>
#include <ctype.h>

int readline (char *curr_line, FILE *fp, char kb[3], short *nlp)
{
    /* Output for readline function:
        data at start of line : 0
        spaces at start of line : 1
        blank line : 2
        first line of record : 3
        end of file : -1
    */
    // declare variables
    char *lp, *test;

    // begin line reading loop
    lp=fgets(curr_line,118,fp);
    if (lp != NULL)
        while ((strlen (curr_line) < 5) && (lp != NULL))
            lp=fgets(curr_line,118,fp);
    *nlp=0;
    // check for errors, EOF condition
    if (lp== NULL) return(-1);
    //test for key at beginning of line
    lp=curr_line;
    lp++;
    kb[0] = *curr_line;
    kb[1] = *lp;
    if (strcmp (kb," ") == 0)
        return(1);
    else if (kb[0]=='\n')
        return(2);
    else if ((test = strstr (curr_line,"##")) != NULL)
        return (3);
    else return(0);
}

int main(int argc, char* argv[])
{
    // set up structure for following data

    // declare variables
    char author[20]; /* author name */
    char *ap; /* pointer to author */
    char line_buffer[120]; /* string for the current line */
    char rec_str[20]; /* string for record identifier */
    char *curr_line; /* pointer to the current line */
    FILE *fp; /* open DIALOG file pointer */
    FILE *ti_fp; /* title file pointer */
    FILE *nr_fp; /* no. refs file pointer */
    FILE *au_fp; /* author file pointer */
    FILE *cs_fp; /* corporate source file pointer */
    FILE *jn_fp; /* journal file pointer */
    FILE *py_fp; /* publication year file pointer */
    FILE *pg_fp; /* page no. file pointer */
    FILE *dt_fp; /* doc type file pointer */
    FILE *de_fp; /* descriptors file pointer */
}
```

⁶¹ Programs co-developed with Andrew Beehag

```

FILE *cr_fp; /* cited reference file pointer */
char *sp, *sp2, *sp3; /* general pointers for line reading */
char tempstr[220], ofname [220]; /* temporary strings */
int li; /* test for current line for presence of key */
// declare variables
char key_buffer[3]; /* string for DIALOG 2-key identifier */
char *pt; /* pointer to key_buffer */
short new_line=0; /* equals 0 when line is new, 1 when line has been processed */
int flag=0, i; /* general flag for string copying */
long refs; /* no refs in NR field */
long upage, lpage; /* upper and lower page numbers */

// open dialog search input file
sp2=sp=tempstr;
if (argv[1] != NULL)
    strcpy (tempstr, argv[1]);
else {
    printf ("Enter the file name      ");
    gets (sp);
}
if ( (fp=fopen(sp,"r")) == NULL ) printf ("File not found\n");

// create string for directory location
while (sp2 != NULL) {
    sp2=sp2+1;
    sp2= strchr(sp, '\\');
}
*sp = '\0';
sp2=tempstr;

pt = key_buffer;
key_buffer[2]='\0';
// read in author surname
ap=author;
printf("Enter the author's name      ");
gets(ap);
sp=ap;
while (*sp != '\0') {
    if (islower(*sp))
        *sp=toupper(*sp);
    sp++;
}

/* _____ */

// find first record
curr_line=line_buffer;
new_line=0;
li=readline(curr_line,fp,key_buffer,&new_line);

// open output files
strcpy (ofname,tempstr);
strcat (ofname,ap);
strcat (ofname,"_ti.csv");
ti_fp=fopen(ofname,"w"); /* title file */

strcpy (ofname,tempstr);
strcat (ofname,ap);
strcat (ofname,"_nr.csv");
nr_fp=fopen(ofname,"w"); /* no. refs file */

strcpy (ofname,tempstr);
strcat (ofname,ap);
strcat (ofname,"_au.csv");
au_fp=fopen(ofname,"w"); /* author file */

strcpy (ofname,tempstr);
strcat (ofname,ap);
strcat (ofname,"_cs.csv");
cs_fp=fopen(ofname,"w"); /* corporate source file */

strcpy (ofname,tempstr);
strcat (ofname,ap);
strcat (ofname,"_jn.csv");
jn_fp=fopen(ofname,"w"); /* journal file */

strcpy (ofname,tempstr);

```

```

    strcat (ofname,ap);
    strcat (ofname,"_py.csv");
    py_fp=fopen(ofname,"w"); /* publn year file */

    strcpy (ofname,tempstr);
    strcat (ofname,ap);
    strcat (ofname,"_pg.csv");
    pg_fp=fopen(ofname,"w"); /* page no file */

    strcpy (ofname,tempstr);
    strcat (ofname,ap);
    strcat (ofname,"_dt.csv");
    dt_fp=fopen(ofname,"w"); /* doc type file */

    strcpy (ofname,tempstr);
    strcat (ofname,ap);
    strcat (ofname,"_de.csv");
    de_fp=fopen(ofname,"w"); /* descriptors file */

    strcpy (ofname,tempstr);
    strcat (ofname,ap);
    strcat (ofname,"_cr.csv");
    cr_fp=fopen(ofname,"w"); /* cited reference file */

// begin loop testing for records and end-of-file
while (li != -1) {

    // case beginning of record
    if (li == 3) {
        sp=strstr(curr_line,"##");
        sp=sp+3;
        strcpy (rec_str,sp);
        sp=strchr (rec_str,'\n');
        *sp='\0';
    }

    // case "NR": place identifier and no refs in *_nr.csv file
    if (strcmp(key_buffer,"NR") == 0) {
        sp=strchr(curr_line,'>');
        sp++;
        while (*sp == ' ') sp++;
        sp2=tempstr;
        while (*sp != '|') {
            *sp2=*sp;
            sp2++;sp++;
        }
        *sp2 = '\0';
        refs = strtol (tempstr,&sp3,10);
        if (refs == 0) {
            li=readline(curr_line,fp,key_buffer,&new_line);
            if ((strcmp(key_buffer,"NR") == 0) &&
(strstr(curr_line,"NO REFS KEYED")) != NULL)
                fprintf (nr_fp,"%s,NO REFS KEYED\n",rec_str);
            else fprintf (nr_fp,"%s,%ld\n",rec_str,refs);
        }
        else fprintf (nr_fp,"%s,%ld\n",rec_str,refs);
    }

}

// case "TI": place identifier and title in *_ti.csv file
if (strcmp(key_buffer,"TI") == 0) {
    flag=0;
    while (flag == 0) { /* flag == 0 if there are more lines to
process */

        /* process line */
        sp=curr_line+4;
        if (*sp==' ') sp++;
        sp2=tempstr;
        while ((*sp != '|') && (flag == 0)) {
            while ((*sp != '|') && (*sp != '\n')) { /* copy
author name */

                *sp2=*sp;
                sp++; sp2++;
                if (*sp == '\n') {

```

```

        li=readline(curr_line,fp,key_buffer,&new_line);
                                sp=curr_line;
                                while (*sp == ' ') sp++;
                                *sp2=' '; sp2++;
                                }
                                }
                                *sp2='\0';
                                fprintf (ti_fp,"%s,%s\n",rec_str,tempstr);
                                sp2=tempstr;
                                if (*sp == '\n') {
        li=readline(curr_line,fp,key_buffer,&new_line);
                                sp=curr_line;
                                while (*sp == ' ') sp++;
                                }
                                /* end process line */
                                if (*sp == '|') flag=1;
                                }
                                }
                                }

//case "AU": find author field, list authors in *_au.csv file
if (strcmp(key_buffer,"AU") == 0) {
    flag=0;
    while (flag == 0) { /* flag == 0 if there are more lines to
process */
        /* process line */
        sp=curr_line+4;
        if (*sp==' ') sp++;
        sp2=tempstr;
        while ((*sp != '|') && (flag == 0)) {
            while ((*sp != '^') && (*sp != '(') && (*sp != '|')
&& (*sp != '\n')) { /* copy author name */
                if (*sp == ',')
                    *sp2=' ';
                else *sp2=*sp;
                sp++; sp2++;
                if (*sp == '\n') {
                    li=readline(curr_line,fp,key_buffer,&new_line);
                                sp=curr_line;
                                while (*sp == ' ') sp++;
                                *sp2=' '; sp2++;
                                }
                                }
                                if (*sp == '(') { /* remove "reprint" from records
*/
                                    while ((*sp != '^') && (*sp != '|') && (*sp
!= '\n')) {
                                        sp++;
                                        if (*sp == '\n') { /* special case
of new line after "reprint" */
                                            li=readline(curr_line,fp,key_buffer,&new_line);
                                                sp=curr_line;
                                                while (*sp == ' ') sp++;
                                                }
                                                }
                                                }
                                                *sp2='\0';
                                                fprintf (au_fp,"%s,%s\n",rec_str,tempstr);
                                                sp2=tempstr;
                                                if (*sp == '^') {
                                                    sp++;
                                                    while (*sp == ' ') sp++;
                                                }
                                                if (*sp == '\n') {
                                                    li=readline(curr_line,fp,key_buffer,&new_line);
                                                        sp=curr_line;
                                                        while (*sp == ' ') sp++;
                                                        }
                                                        }
                                                        /* end process line */
                                                        if (*sp == '|') flag=1;

```

```

    }
}

// case "CS": place identifier and corporate sources one per line in *_cs.csv
file
    if (strcmp(key_buffer,"CS") == 0) {
        flag=0;
        while (flag == 0) { /* flag == 0 if there are more lines to
process */
            /* process line */
            sp=curr_line+4;
            if (*sp=='<') {
                while (*sp!='>') sp++;
                sp++;
            }
            while (*sp==' ') sp++;
            sp2=tempstr;
            while ((*sp != '|') && (flag == 0)) {
                while ((*sp != '^') && (*sp != '|') && (*sp !=
'\n')) { /* copy reference */
                    *sp2=*sp;
                    sp++; sp2++;
                    if (*sp == '\n') {
                        li=readline(curr_line,fp,key_buffer,&new_line);
                        sp=curr_line;
                        while (*sp == ' ') sp++;
                        *sp2=' '; sp2++;
                    }
                }
                if (*sp == '(') { /* remove "reprint" from records
*/
                    while ((*sp != '^') && (*sp != '|') && (*sp
!= '\n')) {
                        sp++;
                        if (*sp == '\n') { /* special case
of new line after "reprint" */
                            li=readline(curr_line,fp,key_buffer,&new_line);
                            sp=curr_line;
                            while (*sp == ' ') sp++;
                        }
                    }
                }
                *sp2='\0';
                /* change commas in address into semicolons */
                sp2=strchr (tempstr,',');
                while (sp2 != NULL) {
                    *sp2 = ';';
                    sp2 = strchr (tempstr,',');
                }
                /* end change commas in address into semicolons */
                fprintf (cs_fp,"%s,%s\n",rec_str,tempstr);
                sp2=tempstr;
                if (*sp == '^') {
                    sp++;
                    while (*sp == ' ') sp++;
                }
                if (*sp == '\n') {
                    li=readline(curr_line,fp,key_buffer,&new_line);
                    sp=curr_line;
                    while (*sp == ' ') sp++;
                }
                if (*sp == '|') flag = 1;
            }
        }
    }

/* case "SO": place identifier and journal title in *_jn.csv file
place identifier and place publication year in
*.py.csv file

```

```

place identifier and place no of pages in *_pg.csv

file */

    if (strcmp(key_buffer,"SO") == 0) {
        sp2=strchr(curr_line,'<');
        sp2++;
        key_buffer[0]=*sp2;
        sp2++;
        key_buffer[1]=*sp2;
        key_buffer[2]='\0';
        if (strcmp(key_buffer,"JN") == 0) {
            sp=strchr(curr_line,'>');
            sp++;
            while (*sp == ' ') sp++;
            sp2=tempstr;
            while ((*sp != '|') && (*sp != '\n')) {
                *sp2=*sp;
                sp++; sp2++;
                if (*sp == '\n') {
                    li=readline(curr_line,fp,key_buffer,&new_line);
                    sp=curr_line;
                    while (*sp == ' ') sp++;
                    *sp2=' '; sp2++;
                }
            }
            *sp2='\0';
            fprintf (jn_fp,"%s,%s\n",rec_str,tempstr);
        }
        else if (strcmp(key_buffer,"PY") == 0) {
            sp=strchr(curr_line,'>');
            sp++;
            while (*sp == ' ') sp++;
            sp2=tempstr;
            while ((*sp != '|') && (*sp != ' ')) {
                *sp2=*sp;
                sp++; sp2++;
                while (*sp == ' ') sp++;
            }
            *sp2='\0';
            fprintf (py_fp,"%s,%s\n",rec_str,tempstr);
        }
        else if (strcmp(key_buffer,"PG") == 0) {
            sp=strchr(curr_line,'>');
            sp++;
            while (*sp == ' ') sp++;
            sp2=tempstr;
            while ((*sp != '|') && (*sp != ' ')) {
                *sp2=*sp;
                sp++; sp2++;
                while (*sp == ' ') sp++;
            }
            *sp2='\0';
            sp3=strchr(tempstr,'-');
            if (sp3 == NULL) fprintf (pg_fp,"%s,1\n",rec_str);
            else {
                *sp3++;
                upage=strtol(sp3,&sp,10);
                sp3=strchr(tempstr,'-');
                *sp3='\0';
                lpage=strtol(tempstr,&sp,10);
                fprintf (pg_fp,"%s,%ld\n",rec_str,(upage-lpage+1));
            }
        }
    }

// case "DT": place identifier and doument type in *_dt.csv file
if (strcmp(key_buffer,"DT") == 0) {
    flag=0;
    while (flag == 0) { /* flag == 0 if there are more lines to
process */
        /* process line */
        sp=curr_line+4;
        if (*sp==' ') sp++;
        sp2=tempstr;

```



```

        while ((*sp != '|') && (flag == 0)) {
            while ((*sp != '|') && (*sp != '\n')) {
                *sp2=*sp;
                sp++; sp2++;
                if (*sp == '\n') {

li=readline(curr_line,fp,key_buffer,&new_line);
                                sp=curr_line;
                                while (*sp == ' ') sp++;
                                *sp2=' '; sp2++;

                                }
                                *sp2='\0';
                                fprintf (dt_fp,"%s,%s\n",rec_str,tempstr);
                                sp2=tempstr;
                                if (*sp == '\n') {

li=readline(curr_line,fp,key_buffer,&new_line);
                                sp=curr_line;
                                while (*sp == ' ') sp++;

                                }
                                /* end process line */
                                if (*sp == '|') flag=1;
                                }
                                }
        }

// case "DE": place identifier and descriptors one per line in *_de.csv file
if (strcmp(key_buffer,"DE") == 0) {
    flag=0;
    while (flag == 0) { /* flag == 0 if there are more lines to
process */
        /* process line */
        sp=strchr(curr_line,'_');
        sp++;
        if (*sp==' ') sp++;
        sp2=tempstr;
        while ((*sp != '|') && (flag == 0)) {
            while ((*sp != ';') && (*sp != '|') && (*sp !=
'\n')) { /* copy author name */
                *sp2=*sp;
                sp++; sp2++;
                if (*sp == '\n') {

li=readline(curr_line,fp,key_buffer,&new_line);
                                sp=curr_line;
                                while (*sp == ' ') sp++;
                                *sp2=' '; sp2++;

                                }
                                }
                                *sp2='\0';
                                fprintf (de_fp,"%s,%s\n",rec_str,tempstr);
                                sp2=tempstr;
                                if (*sp == ';') {
                                    sp++;
                                    while (*sp == ' ') sp++;
                                }
                                if (*sp == '\n') {

li=readline(curr_line,fp,key_buffer,&new_line);
                                sp=curr_line;
                                while (*sp == ' ') sp++;

                                }
                                /* end process line */
                                if (*sp == '|') flag=1;
                                }
                                }
        }
    }

// case "CR": list identifiers and one citation per line in *_cr.csv file
if (strcmp(key_buffer,"CR") == 0) {
    flag=0;
    while (flag == 0) { /* flag == 0 if there are more lines to
process */

```

```

        /* process line */
        sp=curr_line+4;
        if (*sp=='<') {
            while (*sp!='>') sp++;
            sp++;
        }
        if (*sp==' ') sp++;
        sp2=tempstr;
        while ((*sp != '|' ) && (flag == 0)) {
            while ((*sp != '^') && (*sp != '|') && (*sp !=
'\n')) { /* copy reference */
                *sp2=*sp;
                sp++; sp2++;
                if (*sp == '\n') {
                    li=readline(curr_line,fp,key_buffer,&new_line);
                    sp=curr_line;
                    while (*sp == ' ') sp++;
                    *sp2=' '; sp2++;
                }
            }
            *sp2='\0';
        /* remove 3rd and subsequent commas */
        sp3=sp2=tempstr;
        i=0;
        while ((i<3) && (sp3 != NULL)) {
            sp2++;
            sp3= strchr (sp2,',');
            if ((sp3 == NULL) && (i < 2)) *sp2=',';
            else sp2=sp3;
            i++;
        }
        while (sp3 != NULL) {
            sp2=sp3;
            *sp2 = ' ';
            sp3= strchr (sp2,',');
        }
        /* end of remove commas */
        fprintf (cr_fp,"%s,%s\n",rec_str,tempstr);
        sp2=tempstr;
        if (*sp == '^') {
            sp++;
            while (*sp == ' ') sp++;
        }
        if (*sp == '\n') {
            li=readline(curr_line,fp,key_buffer,&new_line);
            sp=curr_line;
            while (*sp == ' ') sp++;
        }
        /* end process line */
        if (*sp == '|') {
            sp++;
            if (*sp == '\n') {
                li=readline(curr_line,fp,key_buffer,&new_line);
                sp=curr_line;
                while (*sp == ' ') sp++;
            }
            if (*sp == '|') flag=1;
            while (*sp == ' ') sp++;
            if (*sp == '\n') { /* special case of new
line after '|' */
                li=readline(curr_line,fp,key_buffer,&new_line);
                sp=curr_line;
                while (*sp == ' ') sp++;
            }
        }
    }
    li=readline(curr_line,fp,key_buffer,&new_line);
}

fclose (ti_fp);

```

```

        fclose (nr_fp);
        fclose (au_fp);
        fclose (cs_fp);
        fclose (jn_fp);
        fclose (py_fp);
        fclose (pg_fp);
        fclose (dt_fp);
        fclose (de_fp);
        fclose (cr_fp);

    return 0;
}

```

APPENDIX B.2. PROGRAM FOR GENERATING A SYMMETRIC MATRIX

```

// Symmetric.cpp : Defines the entry point for the console application.
//

```

```

#include "stdafx.h"
#include <stdio.h>
#include <string.h>
#include <malloc.h>
#include <stdlib.h>
#include <ctype.h>

typedef struct COAUTH {
    short id; /* Author ID */
    short count; /* coauthoring count */
    struct COAUTH *next, *last;
} COAUTHOR;

typedef struct AUTH {
    short id; /* Author ID */
    char name[40]; /* author name */
    short count; /* coauthoring count */
    struct COAUTH *col; /* co-author list */
    struct AUTH *next, *last;
} AUTHOR;

typedef struct ENT {
    short co_id; /* co-author number */
    struct ENT *next, *last;
} JITEM;

int main(int argc, char* argv[])
{
    FILE *infp, *outfp; /* CSV file pointer */
    AUTHOR *first, *current; /* pointers to AUTHOR records */
    JITEM *e_start, *e_curr, *e_co;
    COAUTHOR *co_curr;
    char instr[40], outstr[40]; /* file name strings */
    char tempstr[10];
    char in_line[2000]; /* array for input file line */
    char co_line[2000]; /* listing of coauthor locations */
    char *lp, *co_lp; /* pointer to current line */
    char *lpt; /* test line pointer */
    char *lp_a, *lp_b; /* string manipulation pointers */
    short num_authors; /* total number of authors */
    int i, count=0; /* counting integers */

    // open input file;
    if (argv[1] != NULL){
        infp=fopen(argv[1], "r");
    }
    else {
        infp=NULL;
        while (infp==NULL) {
            printf ("Enter the file name      ");
            gets (instr);
            infp=fopen(instr,"r");
            if (infp==NULL) printf("\nFile not found.\n");
        }
    }
}

```

```

//      read authors line in
      lp=in_line;
      lpt=fgets(lp,2000,infp);

// open output file, write author headings for first line
      printf("\nEnter the output file name      ");
      gets(outstr);
      outfp=fopen(outstr,"w");
      lp_a=strchr(lp,',');
      fprintf(outfp,"%s",lp_a);

// create linked memory structure for authors, including obtaining names
      if (lp!=NULL) {
          printf ("\nCreating memory structure.");
          first=(AUTHOR *)malloc( sizeof( AUTHOR ) );
          current=first;
          first->id=1;
          first->count=0;
          first->col=NULL;
          first->last=NULL;
          first->next=NULL;
          lp_a++;
          lp_b=strchr(lp_a,',');
          *lp_b='\0';
          lpt=strcpy(first->name,lp_a);
          num_authors=1;
          lp_b++;
          lp_a=lp_b;
          lp_b=strchr(lp_a,',');
          while (lp_b!=NULL) {
              *lp_b='\0';
              num_authors++;
              current->next=(AUTHOR *)malloc( sizeof( AUTHOR ) );
              current->next->last=current;
              current=current->next;
              current->id=num_authors;
              lpt=strcpy(current->name,lp_a);
              current->count=0;
              current->col=NULL;
              current->next=NULL;
              lp_b++;
              lp_a=lp_b;
              lp_b=strchr(lp_a,',');
          }
          num_authors++;
          current->next=(AUTHOR *)malloc( sizeof( AUTHOR ) );
          current->next->last=current;
          current=current->next;
          current->id=num_authors;
          if (lp_b=strchr(lp_a,'\n')) *lp_b='\0';
          lpt=strcpy(current->name,lp_a);
          current->count=0;
          current->col=NULL;
          current->next=NULL;
      }

//begin journal list line processing
      lpt=fgets(lp,2000,infp);
      co_lp=co_line;
      *co_lp='\0';
      printf ("\nReading data\n");
      while (lpt!=NULL) {

// obtain list of authors in line
          printf ("#");
          lp_a=strchr(lp,',');
          if (lp_a!=NULL) {
              i=1;
              e_curr=NULL;
              e_start=NULL;
              while (*lp_a!='\0') { /* test for end of line */
                  lp_a++;
                  if
/*
(*lp_a=='1')||(*lp_a=='2')||(*lp_a=='3')||(*lp_a=='4')||(*lp_a=='5')||(*lp_a=='6')||(*lp_a=='7')||(*lp_a=='8')||(*lp_a=='9') {

```

```

        strcat(co_lp, ",");
        lpt=_itoa(i,tempstr,10);
        strcat(co_lp,tempstr);
    }
    original tested version */
    if
    ((*lp_a=='1')||(*lp_a=='2')||(*lp_a=='3')||(*lp_a=='4')||(*lp_a=='5')||(*lp_a=='6')||(*lp_a=='7')||(*lp_a=='8')||(*lp_a=='9')) {
        if (e_curr==NULL) {
            e_curr=(JITEM *)malloc( sizeof( JITEM ) );
            e_start=e_curr;
            e_curr->next=NULL;
            e_curr->last=NULL;
        }
        else {
            e_curr->next=(JITEM *)malloc( sizeof( JITEM
        ) );

            e_curr->next->last=e_curr;
            e_curr=e_curr->next;
            e_curr->next=NULL;
        }
        e_curr->co_id=i;
    }
    while ((*lp_a!='\0')&&(*lp_a!='\0')) lp_a++;
    i++;
}
//
fprintf(outfp,"%s\n",co_lp);

if (e_curr!=NULL) { /* tests to see if entries were present in the
journal */
    e_curr=e_start;
    while (e_curr!=NULL) {
        //
        find first author number (e_curr);
        current=first;
        while (current->id!=e_curr->co_id) current=current-
>next;

        e_co=e_start;
        while (e_co!=NULL) {
            if (current->col==NULL) {
                current->col=(COAUTHOR *)malloc(
sizeof( COAUTHOR ) );

                co_curr=current->col;
                co_curr->last=NULL;
                co_curr->next=NULL;
                co_curr->id=e_co->co_id;
                co_curr->count=1;
            }
            else { /* there are existing entries in
col */
                if (current->col->id>e_co->co_id) {
                    co_curr=current->col;
                    current->col=(COAUTHOR
*)malloc( sizeof( COAUTHOR ) );

                    current->col->next=co_curr;
                    current->col->last=NULL;
                    co_curr->last=current->col;
                    co_curr=co_curr->last;
                    co_curr->id=e_co->co_id;
                    co_curr->count=1;
                }
                else if (current->col->id==e_co-
>co_id)
                    current->col->count++;
                else {
                    co_curr=current->col;
                    while ((co_curr->id<e_co-
>co_id)&&(co_curr->next!=NULL)) co_curr=co_curr->next;

                    if (co_curr->id==e_co-
>co_id)
                        co_curr->count++;
                    else if (co_curr->id>e_co-
>co_id) {
                        /* put new entry in middle of list */
                        co_curr->last-
>next=(COAUTHOR *)malloc( sizeof( COAUTHOR ) );

```

```

>last=co_curr->last;
>next=co_curr;
>last=co_curr->last->next;
>last;
>co_id;

co_curr->last->next-
co_curr->last->next-
co_curr-
co_curr=co_curr-
co_curr->id=e_co-
co_curr->count=1;
}
else if (co_curr-
co_curr-
co_curr->next-
co_curr=co_curr-
co_curr->next=NULL;
co_curr->id=e_co-
co_curr->count=1;
}

}
e_co=e_co->next;
co_curr=current->col; /* ~~ new line entry
*/
}
e_curr=e_curr->next;
}

// free up dynamic memory
e_curr=e_start;
while (e_curr->next!=NULL) {
    e_curr=e_curr->next;
    free(e_curr->last);
}
free(e_curr);
}

// add or append entries for each author listed in the line

        *co_lp='\0';
    }
    lpt=fgets(lp,2000,info);
    count++;
}

// close input file
fclose(info);

// convert data into an output file
printf ("\n\nWriting output file.");
current=first;
co_curr=current->col;
fputs(current->name,outfp);
i=1;
while (co_curr->next!=NULL) {
    while (co_curr->id>i) {
        i++;
        fputc(' ',outfp);
        fputc('0',outfp);
    }
    _itoa(co_curr->count,tempstr,10);
    fputc(' ',outfp);
    fputs(tempstr,outfp);
    i++;
    co_curr=co_curr->next;

```

```

    }
    while (co_curr->id>i) {
        i++;
        fputc(',', outfp);
        fputc('0', outfp);
    }
    _itoa(co_curr->count, tempstr, 10);
    fputc(',', outfp);
    fputs(tempstr, outfp);
    i++;
    while (i<=num_authors) {
        fputc(',', outfp);
        fputc('0', outfp);
        i++;
    }
}

while (current->next!=NULL) {
    fprintf(outfp, "\n");
    current=current->next;
    co_curr=current->col;
    fputs(current->name, outfp);
    i=1;
    if (co_curr==NULL) {
        for (i=1; i<=100; i++) {
            fputc(',', outfp);
            fputc('0', outfp);
        }
    }
    else {
        while ((co_curr!=NULL)&&(co_curr->next!=NULL)) {
            while (co_curr->id>i) {
                i++;
                fputc(',', outfp);
                fputc('0', outfp);
            }
            _itoa(co_curr->count, tempstr, 10);
            fputc(',', outfp);
            fputs(tempstr, outfp);
            i++;
            co_curr=co_curr->next;
        }
        while (co_curr->id>i) {
            i++;
            fputc(',', outfp);
            fputc('0', outfp);
        }
        _itoa(co_curr->count, tempstr, 10);
        fputc(',', outfp);
        fputs(tempstr, outfp);
        i++;
        while (i<=num_authors) {
            fputc(',', outfp);
            fputc('0', outfp);
            i++;
        }
    }
}

// close output file
fclose(outfp);

// release all dynamic memory alloc
current=first;
while (current->next!=NULL) {
    if (current->col!=NULL) {
        co_curr=current->col;
        while (co_curr->next!=NULL) {
            co_curr=co_curr->next;
            free(co_curr->last);
        }
        free(co_curr);
    }
}

```

```
        }
        current=current->next;
        free (current->last);
    }
    free(current);
    printf ("\nDone!\n");

    return 0;
}
/*
```

APPENDIX C. CITING IDENTITY TYPE CLUSTER SOLUTION

APPENDIX C.1. TABLE FOR COMPARISON OF ALTERNATIVE CLUSTER ANALYSIS RESULTS, FOR A 7 CLUSTER SOLUTION

author code	centroid - 7	k-means-7	Ward - 7	Ratio	Gini
NAT	7	4	7	0.1486	0.76743
STM	7	4	7	0.17936	0.71997
SDL	4	3	4	0.22734	0.67557
SHP	4	3	4	0.23148	0.64045
SIW	4	3	4	0.23579	0.63985
RIG	4	3	4	0.24027	0.61653
TEE	4	3	4	0.2433	0.64845
JPN	4	3	4	0.27458	0.61187
STP	4	3	4	0.27558	0.61352
COA	4	3	4	0.27881	0.60389
HTL	4	3	4	0.28391	0.62834
CME	4	3	4	0.29234	0.62996
GJA	4	3	4	0.29262	0.59979
KJF	4	3	4	0.29614	0.6097
SAD	2	5	2	0.30381	0.56173
HJA	4	3	4	0.31198	0.60458
SRN	4	3	4	0.31308	0.5929
SHT	4	3	4	0.3149	0.59083
LIB	4	3	4	0.31633	0.61873
RDM	2	5	2	0.31754	0.5655
KOP	2	5	2	0.31944	0.54565
EJC	2	5	2	0.321	0.57099
KOC	2	5	2	0.32469	0.53742
DRJ	2	5	2	0.33262	0.57907
TJG	2	5	2	0.33744	0.57129
GSA	2	5	2	0.33969	0.57303
SRW	2	5	2	0.34143	0.54869
MAD	2	5	2	0.34309	0.54445
EGM	2	5	2	0.34347	0.58188
BBJ	2	5	2	0.34381	0.56866
GMS	2	5	2	0.34732	0.56382
RDM	2	5	2	0.3474	0.56992
SAC	2	5	2	0.35138	0.54223
WEL	2	5	2	0.35491	0.57365
GMA	2	5	2	0.35518	0.5411
JRS	2	5	2	0.35535	0.5652
TCT	2	5	2	0.36623	0.57189
GCM	2	5	2	0.36625	0.50945
PRJ	2	5	2	0.3689	0.54687
BDJ	2	5	2	0.36898	0.54937
BLP	2	5	2	0.36969	0.53263
TJA	2	5	2	0.37024	0.53449
SGE	2	5	2	0.37137	0.54782
LRR	2	5	2	0.37432	0.54765

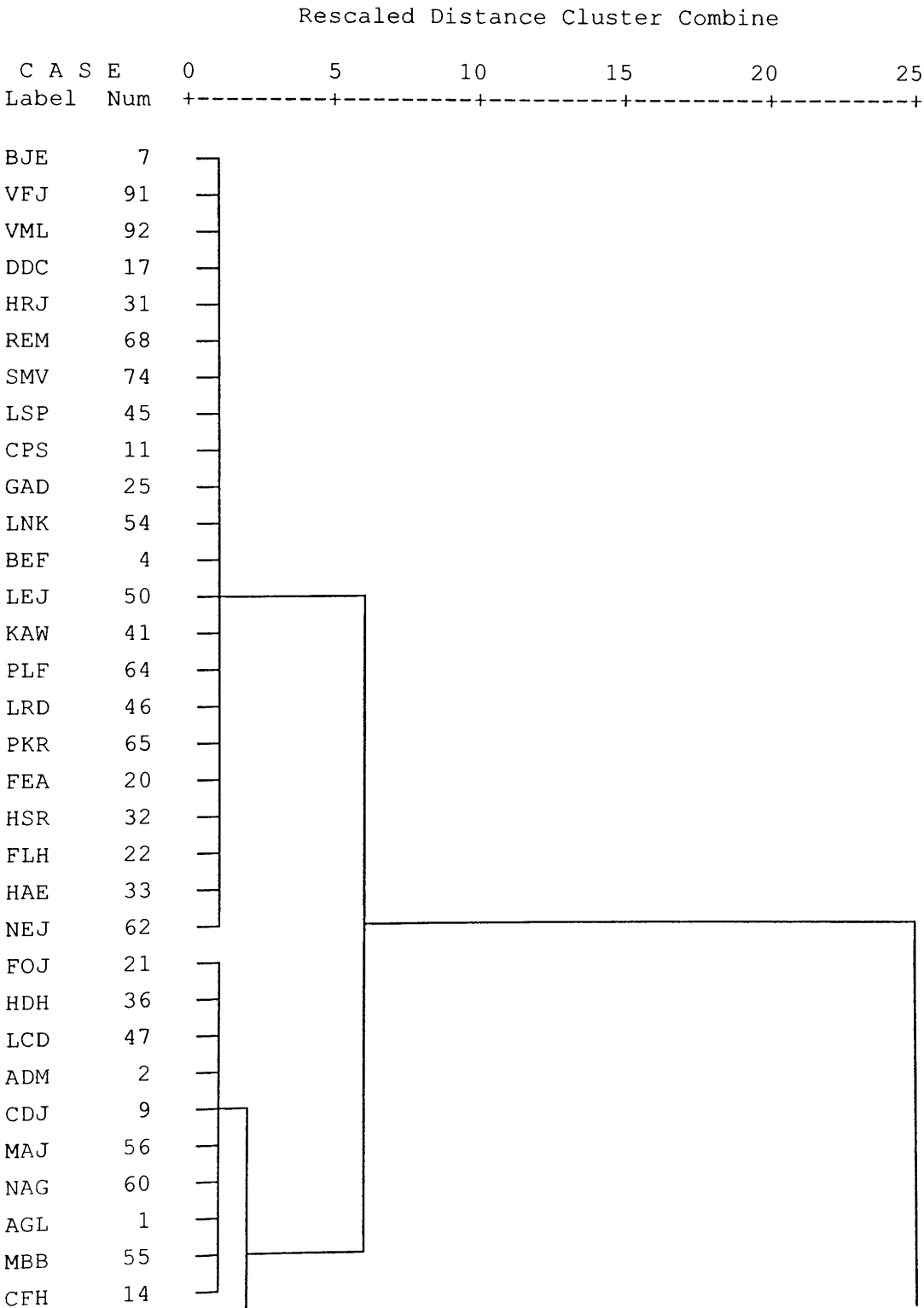
DAR	2	5	2	0.37949	0.5378
BNJ	2	5	2	0.3801	0.51165
FRH	2	5	2	0.38014	0.54566
PMI	2	5	2	0.38243	0.5291
WRN	2	5	2	0.38277	0.53886
PKH	2	5	2	0.38435	0.52304
JIM	2	5	2	0.38647	0.51335
LDS	2	5	2	0.38658	0.52871
LEJ	3	2	3	0.3907	0.449
CPM	2	5	2	0.39112	0.54273
ROY	2	5	2	0.39485	0.51966
SJR	2	5	2	0.39663	0.54289
ZAE	2	5	2	0.39961	0.51441
FNP	2	5	2	0.40259	0.49789
MOG	2	5	2	0.4058	0.49293
SEJ	2	5	2	0.4125	0.50909
LNK	3	2	3	0.42163	0.46556
GAD	3	2	3	0.42962	0.46104
BEF	3	2	3	0.44065	0.4641
VML	3	2	3	0.456	0.4588
BJE	3	2	3	0.45795	0.4616
VFJ	3	2	3	0.45822	0.46193
DDC	3	2	3	0.46521	0.48252
HRJ	3	2	3	0.466	0.46955
SMV	3	2	3	0.46677	0.44585
REM	3	2	3	0.46901	0.44555
LSP	3	2	3	0.47644	0.43361
CPS	3	2	3	0.48355	0.45706
HAE	3	2	3	0.48746	0.42708
NEJ	3	2	3	0.48852	0.39828
HSR	3	2	3	0.49066	0.43235
FEA	3	2	3	0.49441	0.43437
FLH	3	2	3	0.49516	0.42215
KAW	3	2	3	0.50603	0.41639
PLF	3	2	3	0.51563	0.41564
LRD	3	2	3	0.52062	0.40632
PKR	3	2	3	0.52685	0.43299
ADM	1	7	1	0.54938	0.402
CDJ	1	7	1	0.552	0.38307
LCD	1	7	1	0.5613	0.37648
HDH	1	7	1	0.56944	0.37127
FOJ	1	7	1	0.57086	0.37505
MAJ	1	7	1	0.58401	0.3493
NAG	1	7	1	0.59259	0.36008
MBB	1	7	1	0.60606	0.33889
AGL	1	7	1	0.61473	0.3389
CFH	1	7	1	0.62366	0.30953
LMJ	5	1	5	0.66176	0.28529
LDA	5	1	5	0.67335	0.29229
HDO	5	1	5	0.67391	0.31136
WFA	5	1	5	0.69231	0.26401
WRC	6	1	6	0.75431	0.21197

MIN	6	6	6	0.83333	0.15226
-----	---	---	---	---------	---------

APPENDIX C.2. 7 CLUSTER SOLUTION USING WARD METHOD OF CLUSTER ANALYSIS

* * 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 Ward Method



LDA	48	┌
LMJ	53	├
HDO	34	├
WFA	96	├
MIN	57	├
WRC	94	├
KOC	43	├
KOP	44	├
EJC	18	├
RDM	70	├
SAD	73	├
GMA	27	├
SAC	77	├
MAD	58	├
SRW	82	├
JRS	38	├
WEL	95	├
TCT	87	├
BBJ	3	├
RDM	72	├
GMS	26	├
GSA	30	├
TJG	88	├
DRJ	16	├
EGM	19	├
CPM	10	├
SJR	78	├
BDJ	8	├
PRJ	63	├
SGE	76	├
LRR	52	├
BLP	6	├
TJA	90	├
DAR	15	├
WRN	93	├
FRH	24	├
FNP	23	├
MOG	59	├
SEJ	83	├
BNJ	5	├
JIM	39	├
GCM	28	├
LDS	49	├
PMI	66	├
PKH	67	├
ROY	71	├

ZAE	97	└─┘
NAT	61	└─┘
STM	85	└─┘
SIW	81	└─┘
SHP	84	└─┘
TEE	89	└─┘
RIG	69	└─┘
SDL	75	└─┘
SHT	79	└─┘
SRN	80	└─┘
HJA	35	└─┘
LIB	51	└─┘
CME	12	└─┘
HTL	37	└─┘
GJA	29	└─┘
KJF	42	└─┘
JPN	40	└─┘
STP	86	└─┘
COA	13	└─┘

APPENDIX D. DETAILS FROM REPRESENTATIONAL SPACE AND CITING IDENTITY ANALYSIS (CHAPTER 5)

APPENDIX D.1. AUTHORS THAT CO-AUTHOR BUT NOT WITHIN CONSCIOUSNESS PARTICIPANT SET

Authors (17) that have co-authored but do not share any of their co-authors with other consciousness authors, they therefore do not appear in the co-author matrix, this leaves 76 authors in consciousness co-author set

- Armstrong – coauthors with 4 co-authors once each – none of the other authors co-author with the same co-authors
- Block – 3 co-authors once each
- Conrad – 34 co-authors, with which publish 57 times – but none in set
- Fenigstein – 8 co-authors publish with once each
- Flohr – 128 co-authorships with 54 co-authors – none in set though
- Galin – 90 co-authorships with 27 co-authors – none shared
- Jackendorff – 23 co-authorships with 13 co-authors – but none shared
- Laughlin – 10 co-authorships with 8 co-authors – none shared
- Lockwood – 5 co-authorships with 4 co-authors – none shared
- Mangan – 1 co-authorship with 1 co-author – not shared
- Marcel – 7 co-authorships with 5 co-authors – none shared
- Moruzzi, - Moruzzi has co-authors 5 times, with 4 co-authors but drops out of symmetrical matrix as none of his co-authors co-author with any other consciousness authors
- Nagel – 6 co-authorships with 6 co-authors – none shared
- Popper – 29 co-authorships with 26 co-authors – none shared
- Tart – 26 co-authorships with 20 co-authors – none shared
- Velmans – 9 co-authorships with 9 co-authors – none shared
- Wolf – 2 co-authorships with 2 co-authors – none shared

APPENDIX D.2. JOURNALS THAT ARE PUBLISHED ACROSS ALL JOURNAL CLUSTERS

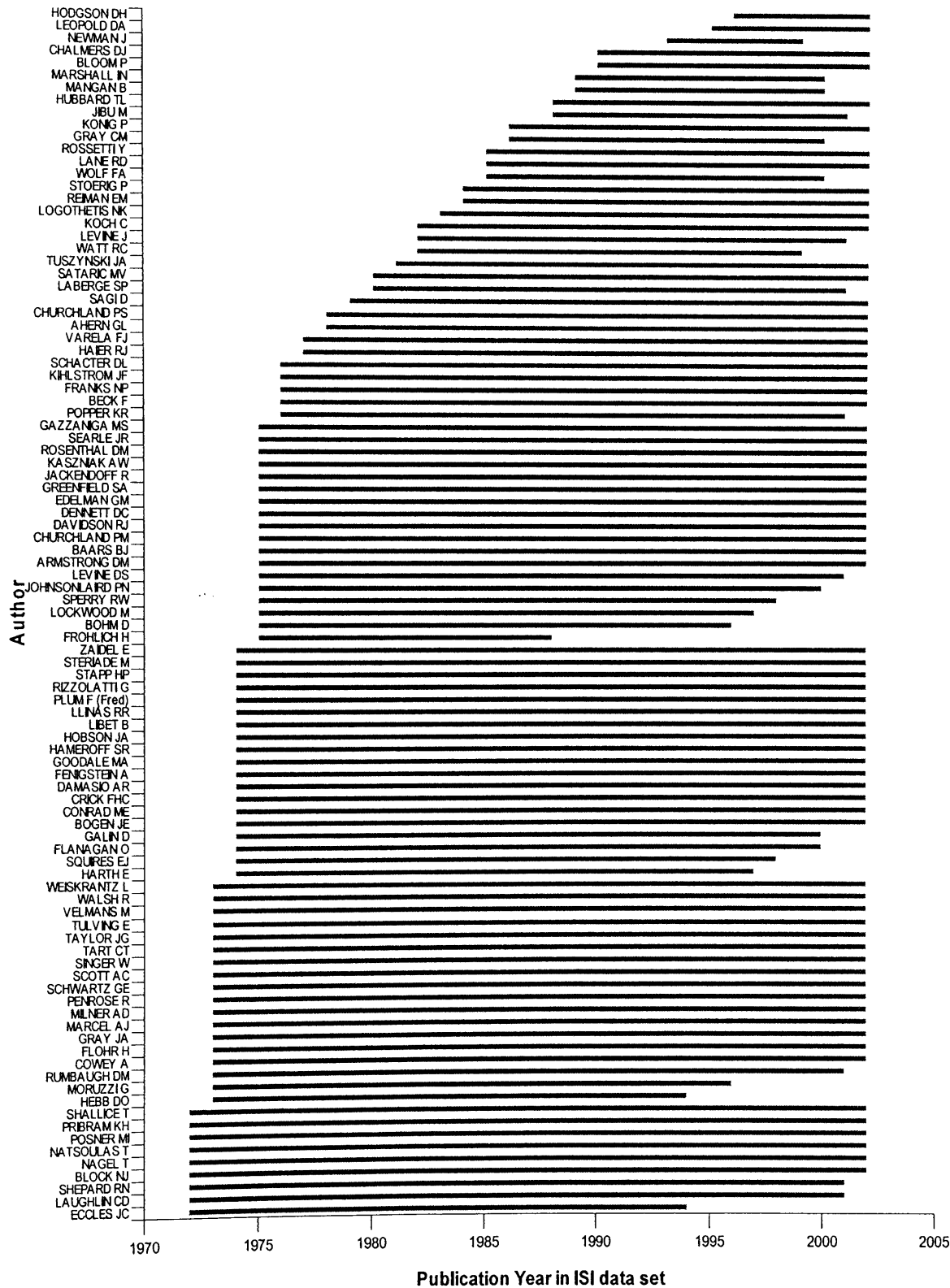
- Annals of the New York Academy of sciences
- Behavioral and brain sciences
- Nature
- Proceedings of the National Academy of Sciences of the United States of America

APPENDIX D.3. JOURNALS THAT ARE PUBLISHED IN BY 6 JOURNAL CLUSTERS

- Biological Psychiatry
- Brain
- Brain research
- Bulletin of the British Psychological Society
- Bulletin of the Psychonomic Society
- Consciousness and cognition
- Contemporary Psychology
- Cortex
- Electroencephalography and clinical neurophysiology
- Experimental brain research
- International journal of neuroscience
- International journal of psychology
- International journal of psychophysiology
- Journal of cognitive neuroscience

-
- Journal of consciousness studies
 - Neuropsychologia
 - New scientist
 - Perception and psychophysics
 - Philosophical transactions of the Royal Society of London series B – biological sciences
 - Science
 - Trends in cognitive sciences
 - Trends in neurosciences

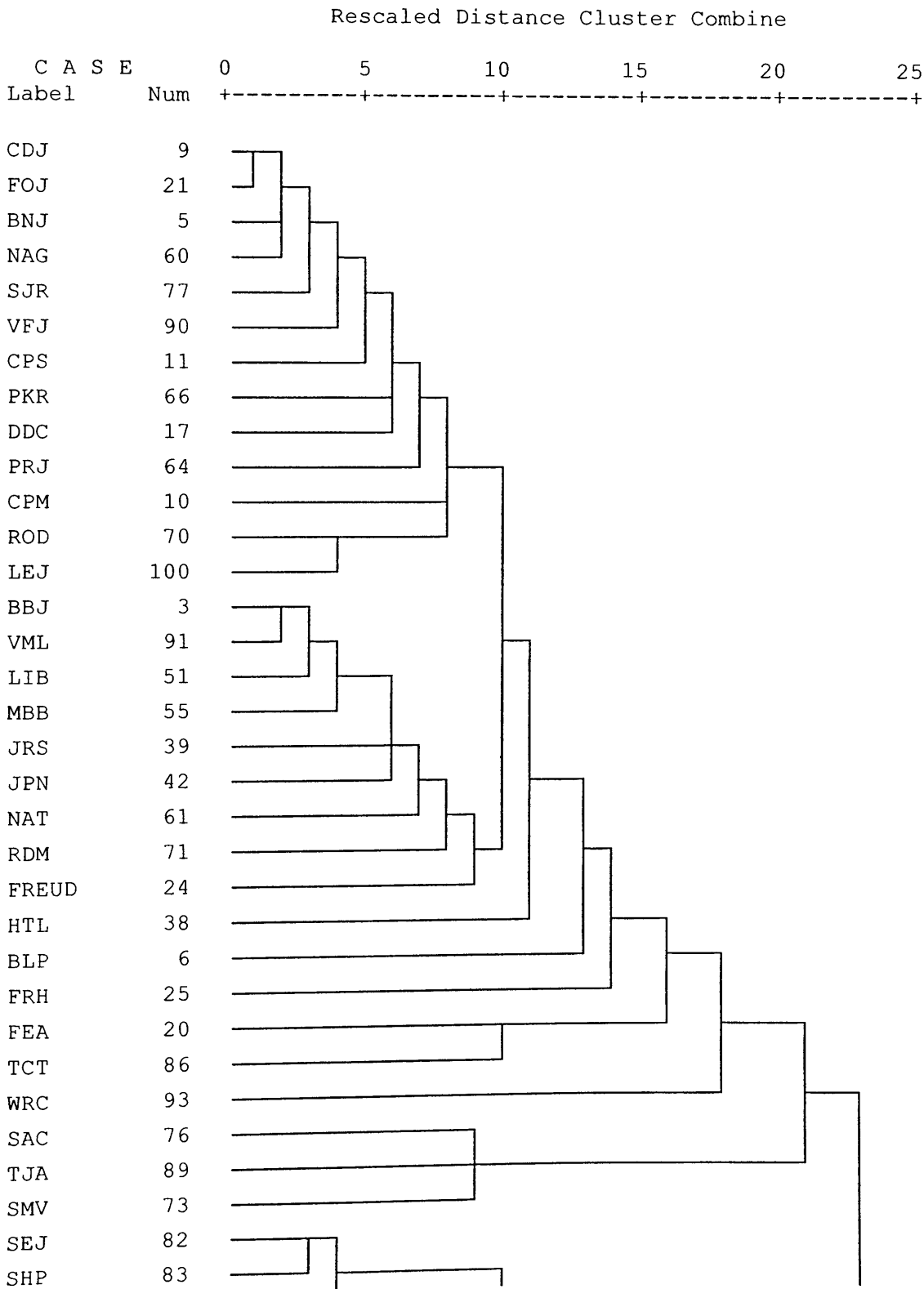
APPENDIX D.4. PUBLICATION CAREER YEAR RANGE FOR INDIVIDUAL AUTHORS

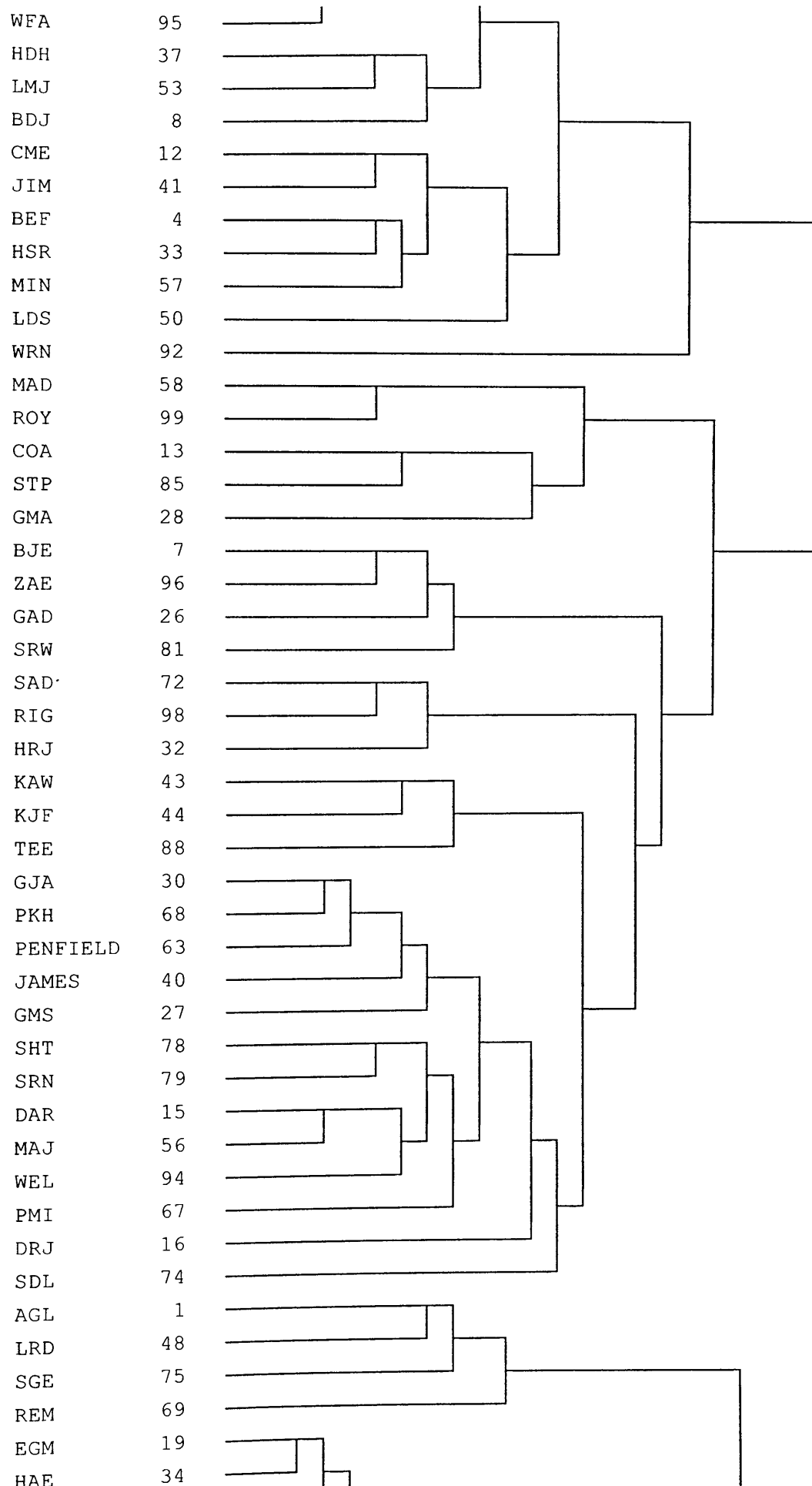


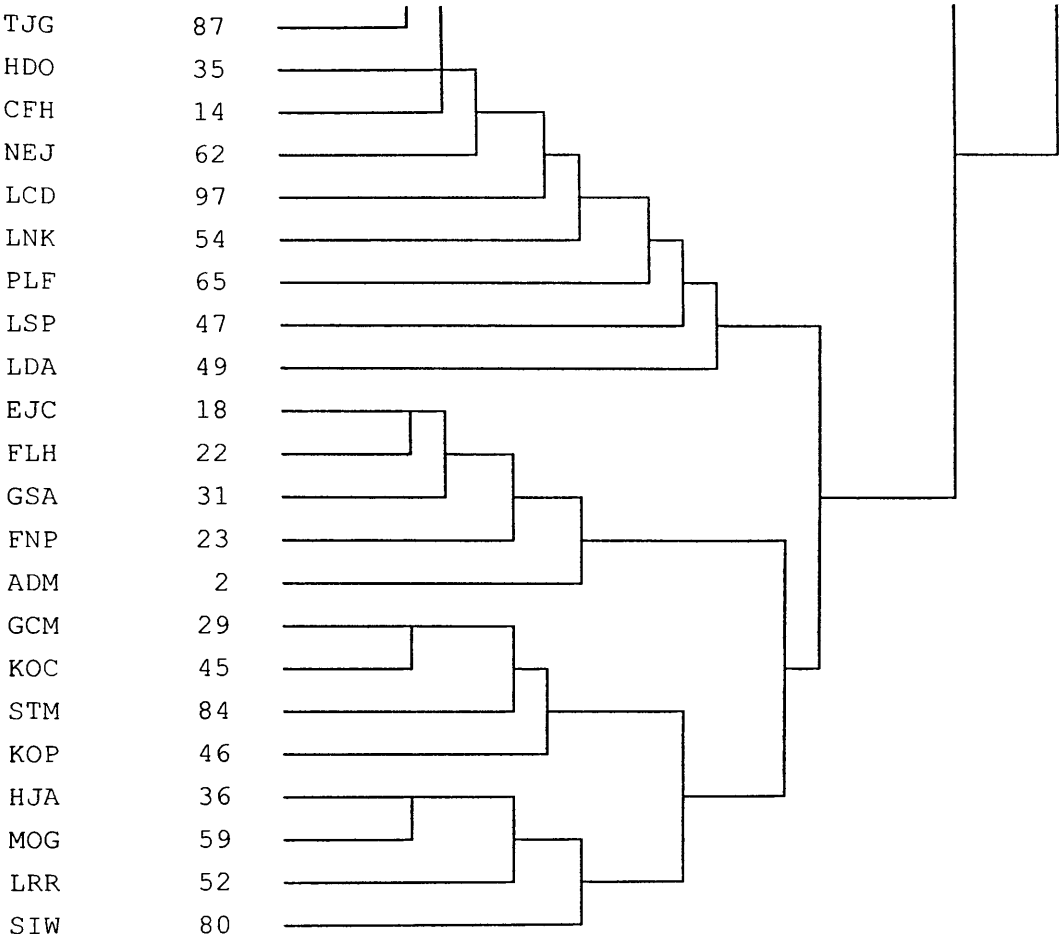
APPENDIX E. CO-AUTHOR CITATION MATRIX CLUSTER SOLUTION

* * * 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 Average Linkage (Within Group)





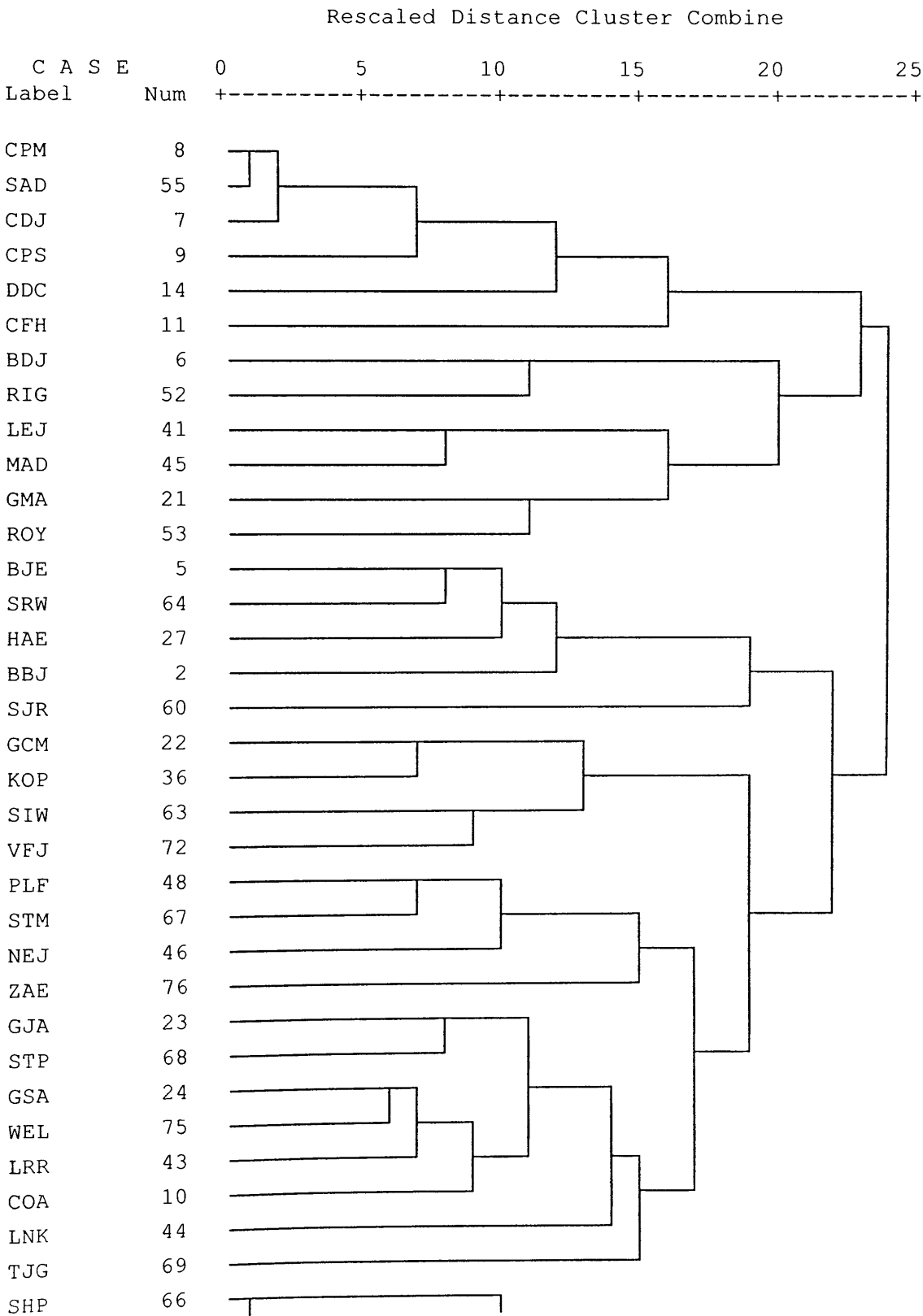


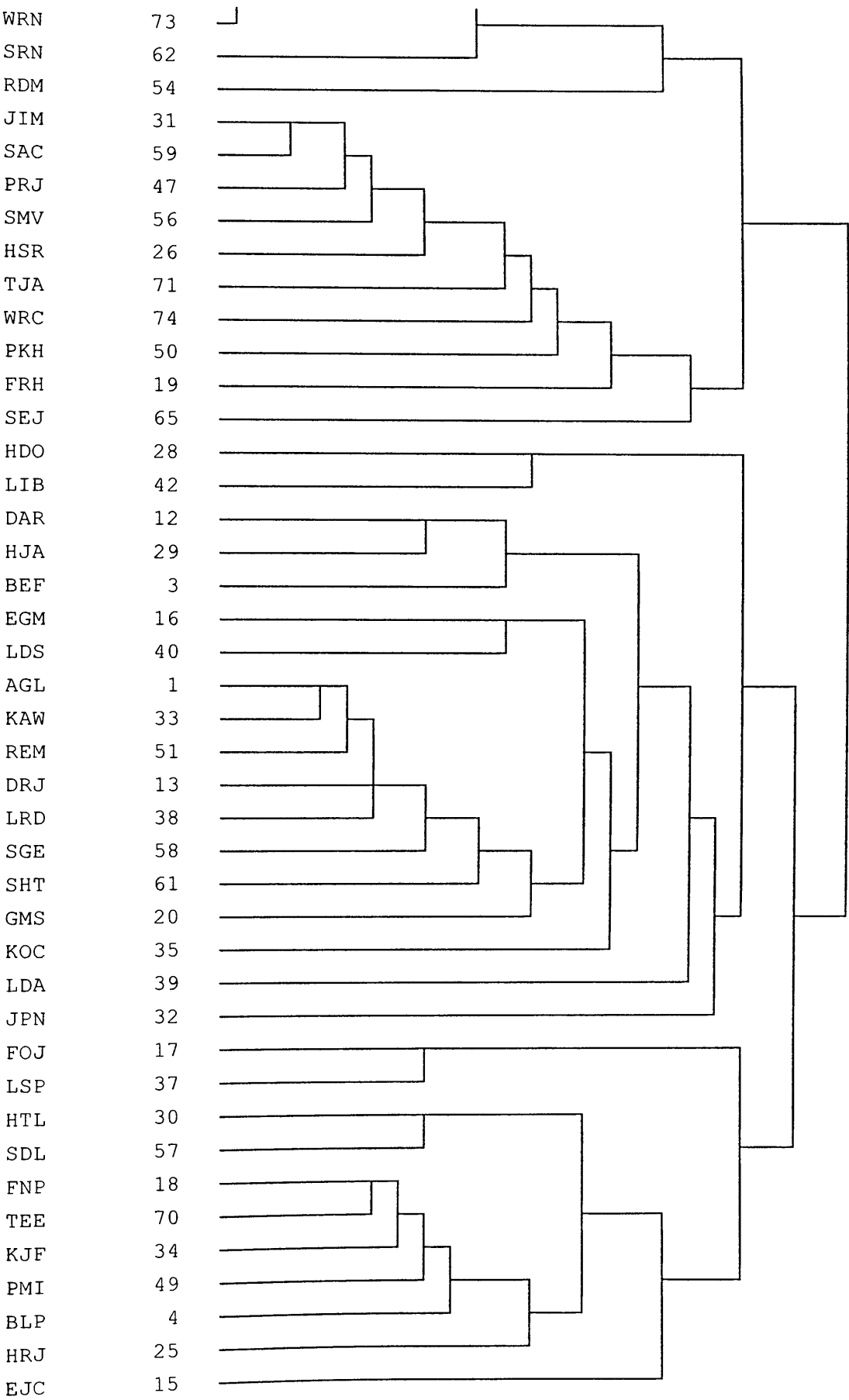
APPENDIX F. CO-AUTHOR MATRIX CLUSTER SOLUTION

Coauthor – symmetrical matrix – authors that share more than one co-author included only

* * 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 Average Linkage (Within Group) --- Cosine

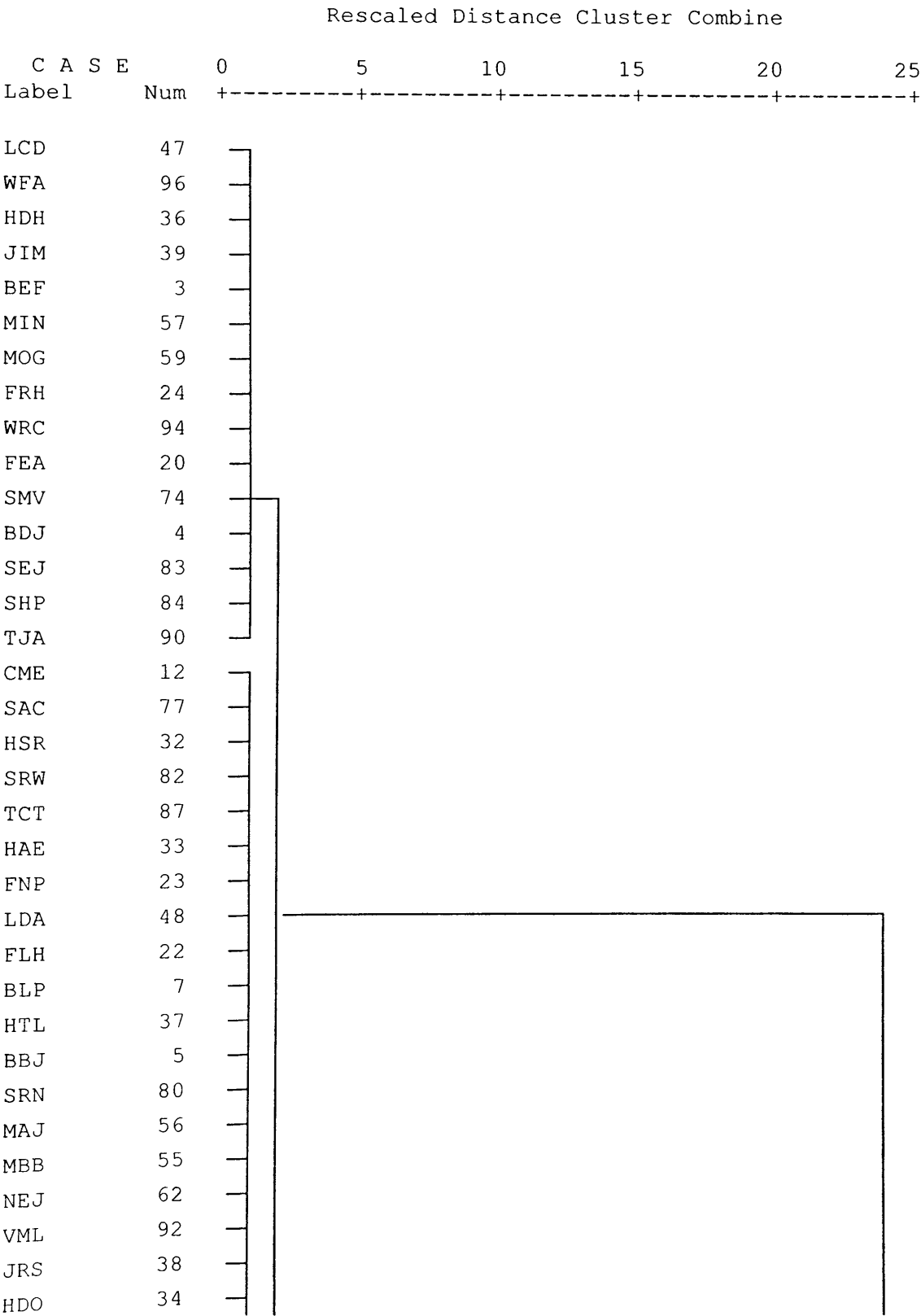




APPENDIX G. CO-JOURNAL PUBLICATION MATRIX
CLUSTER SOLUTION

* * * 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 Ward Method



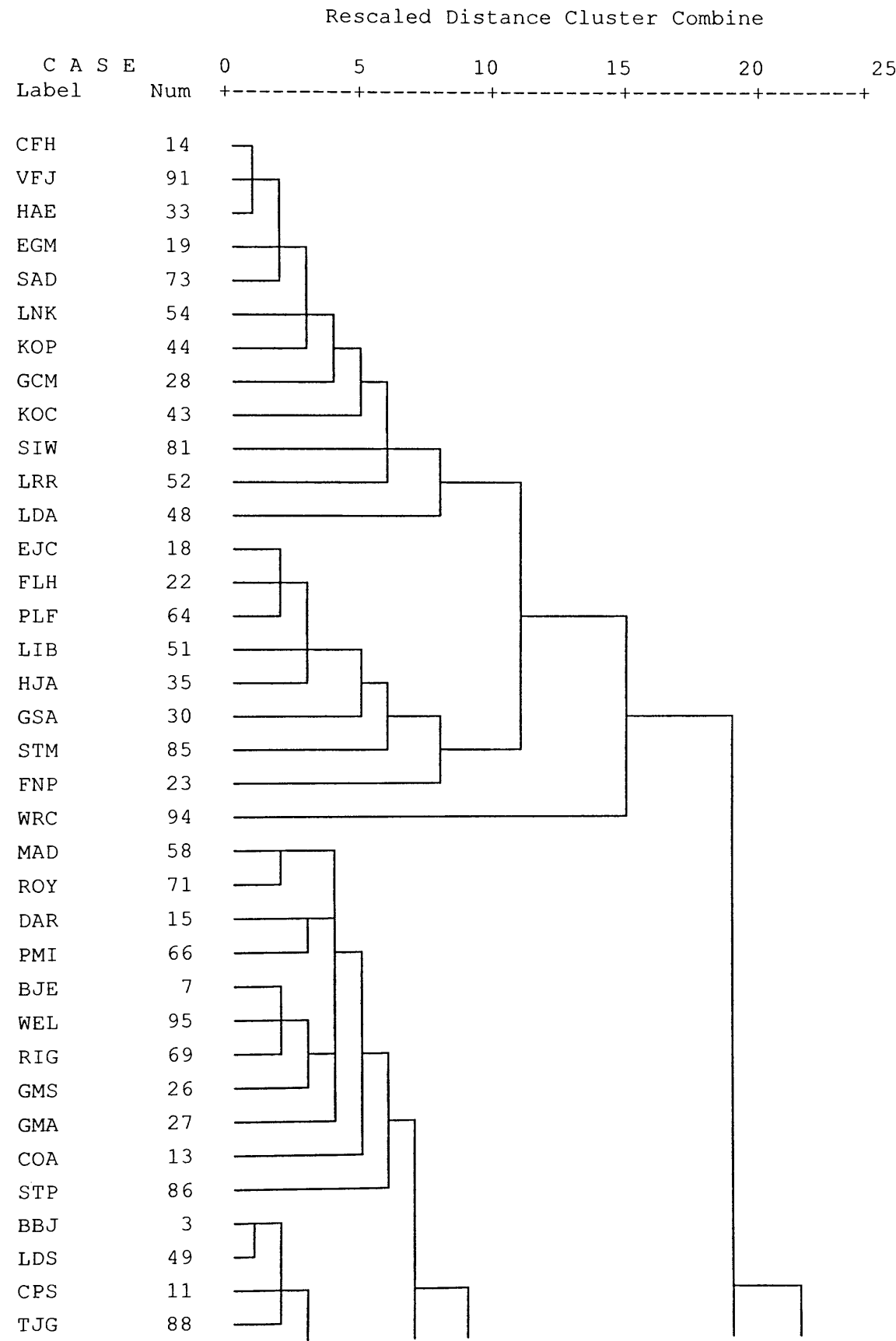
NAT	61	└─┘	
LSP	45	└─┘	
LDS	49	└─┘	
WRN	93	└─┘	
CDJ	9	└─┘	
ROD	70	└─┘	
BNJ	6	└─┘	
FOJ	21	└─┘	
CPM	10	└─┘	
LEJ	50	└─┘	
NAG	60	└─┘	
LMJ	53	└─┘	
ADM	2	└─┘	
PRJ	63	└─┘	
PKR	65	└─┘	
CPS	11	└─┘	
SJR	78	└─┘	
DDC	17	└─┘	
GMS	26	└─┘	
GJA	29	└─┘	
DAR	15	└─┘	
PMI	66	└─┘	
SDL	75	└─┘	
COA	13	└─┘	
LRR	52	└─┘	
LNK	54	└─┘	
STM	85	└─┘	
EGM	19	└─┘	
KOC	43	└─┘	
GMA	27	└─┘	
WEL	95	└─┘	
HJA	35	└─┘	
VFJ	91	└─┘	
MAD	58	└─┘	
RIG	69	└─┘	
SIW	81	└─┘	
GSA	30	└─┘	
LIB	51	└─┘	
CFH	14	└─┘	
EJC	18	└─┘	
GCM	28	└─┘	
SAD	73	└─┘	
KOP	44	└─┘	
ROY	71	└─┘	
STP	86	└─┘	

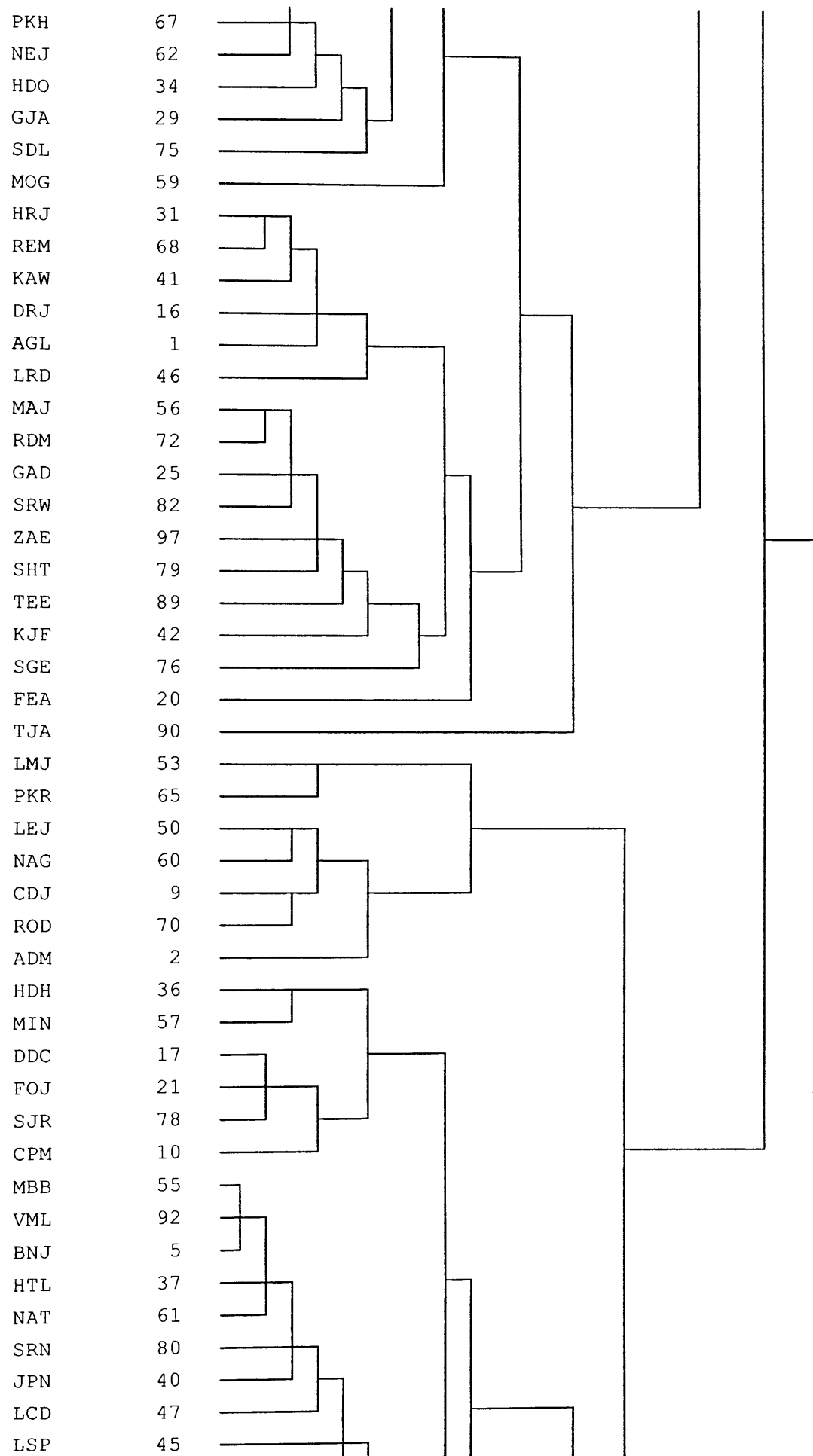
TJG	88	└─┘	
REM	68	└─┘	└──────────┘
ZAE	97	└─┘	
SGE	76	└─┘	
DRJ	16	└─┘	
SHT	79	└─┘	
TEE	89	└─┘	
GAD	25	└─┘	
LRD	46	└─┘	
AGL	1	└─┘	
KAW	41	└─┘	
PLF	64	└─┘	
JPN	40	└─┘	
KJF	42	└─┘	
BJE	8	└─┘	
HRJ	31	└─┘	
RDM	72	└─┘	
PKH	67	└─┘	

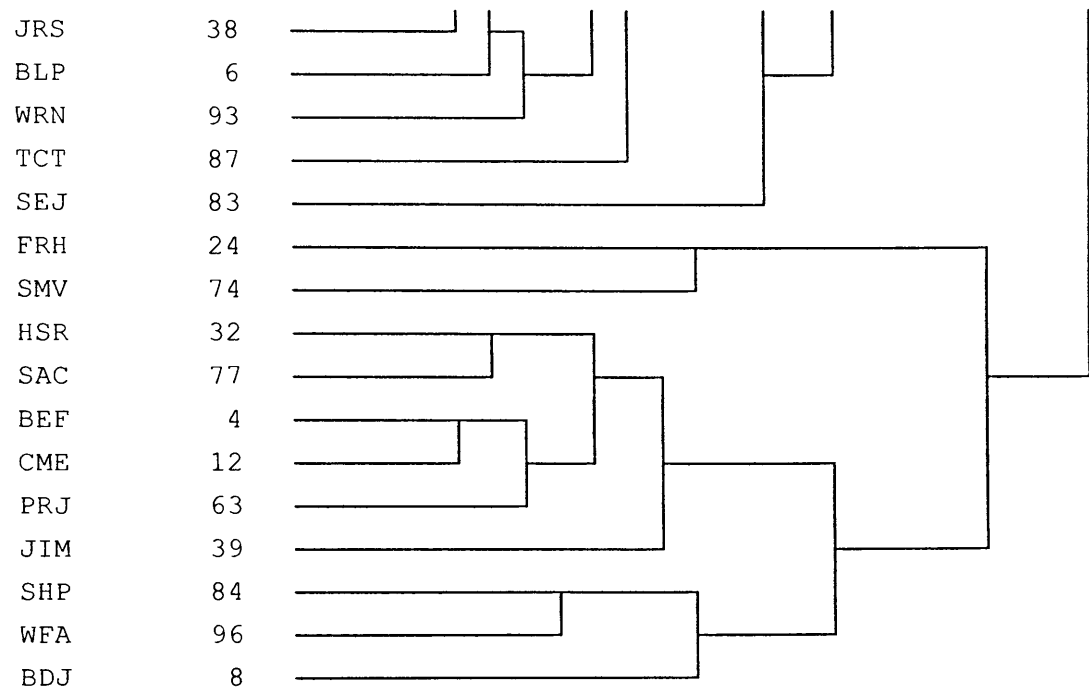
APPENDIX H. CO-CITEE MATRIX CLUSTER SOLUTION

* * * 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 Average Linkage (Within Group) - Cosine









www.thesisonline.com.au

9212 4144

Shop 1, 180 Broadway BROADWAY
702 Harris St ULTIMO 9280 4244

