THE COMBINED APPLICATION OF BIBLIOGRAPHIC COUPLING AND THE COMPLETE LINK CLUSTER METHOD IN BIBLIOMETRIC SCIENCE MAPPING

BO JARNEVING

VALFRID

THE COMBINED APPLICATION OF BIBLIOGRAPHIC COUPLING AND THE COMPLETE LINK CLUSTER METHOD IN BIBLIOMETRIC SCIENCE MAPPING

BO JARNEVING

Akademisk avhandling som med tillstånd av samhällsvetenskapliga fakulteten vid Göteborgs universitet för vinnande av doktorsexamen framläggs till offentlig granskning kl 13.15 fredagen den 10 februari 2006 i Stora hörsalen (C203), Högskolan i Borås, Allégatan 1, Borås.

Institutionen Biblioteks- och informationsvetenskap/Bibliotekshögskolan
Högskolan i Borås och Göteborgs universitet

Title: **The combined application of bibliographic coupling and the complete link cluster method in bibliometric science mapping**

Abstract:
This thesis connects to previous research in bibliometric science mapping and citation indexing. A method was suggested for science mapping purposes and evaluated. The suggestion of this method was motivated by the fact that the prevailing method of citation based science mapping of documents, the cocitation cluster analytical method, can not map the most current published research, a feature that is a characteristic of the proposed method. On theoretical grounds, it was assumed that neither of these methods could substitute for the other and that they would have complementary functions in relation to one another.

The prime objective was to evaluate the proposed method's capability to generate subject coherent clusters, i.e. to identify coherent research themes, and the assumed context of application was scientific information provision. The proposed method has two primary components: (1) a measure of document similarity, bibliographic coupling and (2) a cluster analytical method for the partition of document populations, the complete link cluster method.

The research design comprised four different research settings of which three correspond to specific fields of research and one to a large multidisciplinary environment. Methods of evaluation comprised quantitative approaches as well as more qualitative ones. For the establishment of cluster coherence, measures of density and average coupling strength in clusters were applied. The relevance of generated clusters was assumed to be reflected by these measures and was substantiated by field experts' evaluations of clustering results. In order to assess the agreement between field experts' apprehensions of their fields' cognitive structures, intellectual-manual partitions of document populations were performed by field experts and compared with partitions generated by the proposed method.

Findings showed that the proposed method has the capability to identify and map current and coherent research themes on the level of a single research field as well as in a multidisciplinary environment. However, based on theoretical considerations as well as on empirical findings, it was concluded that it would not suffice as a standard science mapping method where exhaustive depictions of specialties' cognitive structures are aimed at. The reasons for this were:

i. As for now, the method of bibliographic coupling can not identify the most central concepts of a research specialty.

ii. The dependency of consensual referencing implies that only minor shares of original document populations will be available for analysis.

iii. The lack of a method for the decision of appropriate thresholds of coupling strength implies arbitrary threshold settings.

iv. The partition of document populations brought about a fragmentation of research fields.

v. Partitions generated by field experts deviated considerably from partitions generated by the complete link cluster method.

It was therefore concluded that the proposed method may be complementary to the cocitation cluster analytical method and to traditional citation indexing. Based on the empirical findings, a tentative outline for such an application was given.

Keywords: bibliometrics, bibliographic coupling, science mapping, citation indexing, cocitation analysis, cluster analysis, scientific information provision

THE COMBINED APPLICATION OF BIBLIOGRAPHIC COUPLING AND THE COMPLETE LINK CLUSTER METHOD IN BIBLIOMETRIC SCIENCE MAPPING

BO JARNEVING

Akademisk avhandling som med tillstånd av samhällsvetenskapliga fakulteten vid Göteborgs universitet för vinnande av doktorsexamen framläggs till offentlig granskning kl 13.15 fredagen den 10 februari 2006 i Stora hörsalen (C203), Högskolan i Borås, Allégatan 1, Borås.

Institutionen Biblioteks- och informationsvetenskap/Bibliotekshögskolan
Högskolan i Borås och Göteborgs universitet

Title: **The combined application of bibliographic coupling and the complete link cluster method in bibliometric science mapping**

Abstract:
This thesis connects to previous research in bibliometric science mapping and citation indexing. A method was suggested for science mapping purposes and evaluated. The suggestion of this method was motivated by the fact that the prevailing method of citation based science mapping of documents, the cocitation cluster analytical method, can not map the most current published research, a feature that is a characteristic of the proposed method. On theoretical grounds, it was assumed that neither of these methods could substitute for the other and that they would have complementary functions in relation to one another.

The prime objective was to evaluate the proposed method's capability to generate subject coherent clusters, i.e. to identify coherent research themes, and the assumed context of application was scientific information provision. The proposed method has two primary components: (1) a measure of document similarity, bibliographic coupling and (2) a cluster analytical method for the partition of document populations, the complete link cluster method.

The research design comprised four different research settings of which three correspond to specific fields of research and one to a large multidisciplinary environment. Methods of evaluation comprised quantitative approaches as well as more qualitative ones. For the establishment of cluster coherence, measures of density and average coupling strength in clusters were applied. The relevance of generated clusters was assumed to be reflected by these measures and was substantiated by field experts' evaluations of clustering results. In order to assess the agreement between field experts' apprehensions of their fields' cognitive structures, intellectual-manual partitions of document populations were performed by field experts and compared with partitions generated by the proposed method.

Findings showed that the proposed method has the capability to identify and map current and coherent research themes on the level of a single research field as well as in a multidisciplinary environment. However, based on theoretical considerations as well as on empirical findings, it was concluded that it would not suffice as a standard science mapping method where exhaustive depictions of specialties' cognitive structures are aimed at. The reasons for this were:

i. As for now, the method of bibliographic coupling can not identify the most central concepts of a research specialty.

ii. The dependency of consensual referencing implies that only minor shares of original document populations will be available for analysis.

iii. The lack of a method for the decision of appropriate thresholds of coupling strength implies arbitrary threshold settings.

iv. The partition of document populations brought about a fragmentation of research fields.

v. Partitions generated by field experts deviated considerably from partitions generated by the complete link cluster method.

It was therefore concluded that the proposed method may be complementary to the cocitation cluster analytical method and to traditional citation indexing. Based on the empirical findings, a tentative outline for such an application was given.

Keywords: bibliometrics, bibliographic coupling, science mapping, citation indexing, cocitation analysis, cluster analysis, scientific information provision

THE COMBINED APPLICATION OF BIBLIOGRAPHIC COUPLING AND THE COMPLETE LINK CLUSTER METHOD IN BIBLIOMETRIC SCIENCE MAPPING

BO JARNEVING

VALFRID 2005

THE COMBINED APPLICATION OF BIBLIOGRAPHIC COUPLING AND THE COMPLETE LINK CLUSTER METHOD IN BIBLIOMETRIC SCIENCE MAPPING

BO JARNEVING

DOCTORAL THESIS
DEPARTMENT OF LIBRARY AND INFORMATION SCIENCE/SWEDISH SCHOOL OF LIBRARY AND INFORMATION SCIENCE

UNIVERSITY COLLEGE OF BORÅS/GÖTEBORG UNIVERSITY

THE COMBINED APPLICATION OF BIBLIOGRAPHIC COUPLING AND THE COMPLETE LINK CLUSTER METHOD IN BIBLIOMETRIC SCIENCE MAPPING

BO JARNEVING

Distribution:
The Publishing Association Valfrid
Department of Library and Information Science/Swedish School of Library and Information Science
University College of Borås/Göteborg University

Print:
Intellecta Docusys, 2005

Series:
Publications from Valfrid, nr 30

ISBN 91-89416-12-0.
ISSN 1103-6990

ACKNOWLEDGEMENTS

I wish to thank Olle Persson and Elena Macevičiūtė for their supervision of this thesis.

I also wish to express my gratitude to Per Ahlgren for the many fruitful discussions and good advice and to Ronald Rousseau for his good advice and suggestions. Many thanks also to Johan Eklund for providing me with the much needed technical support and for programming.

Further, thanks to Boel Bissmarck for checking the English and to Christian Swalander for editing.

I would also like to thank the following researchers for taking time off their busy schedules to evaluate the relevance of the mapping results of the empirical studies:

- Bengt Alrud
- Kim Bolton
- Richard Danell
- Anders Kastberg
- Göran Levan
- Peder Svensson

Länghem, 2005

Bo Jarneving

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Bibliometrics is the quantitative study of patterns derived from the production and use of publications. It was defined by Pritchard in 1969 as "the application of mathematical and statistical methods to books and other media of communication". It is most often used in the field of library and information science, but has also wide applications in other areas (e.g. science policy).

An important area of bibliometric research is *citation analysis*. This sub-field comprises several methods for the analysis of citation relations in research literatures. The analysis of citations originates from the need of scientists to build on previous research when embarking on new research projects and to refer back to them when publishing the results. When referring back to previous research, the publishing scientist sets the framework of his research, while the publishing of the research itself can be seen as the individual scientist's claim of intellectual property and the seeking for acknowledgement by peers. This acknowledgement is in turn reflected by possible future citations in other scientists' subsequent publications. In Ziman (1984, p. 58) it is stated that:

> ...the basic principle of academic science is that results of research must be made public /.../. Whatever scientists think or say individually, their discoveries cannot be regarded as belonging to scientific knowledge until they have been reported to the world and put on permanent record.

Based on the needs of scientists to find and reference previous published research, so called *citation indexes* have been constructed. A citation index facilitates the retrieval of documents associated through citation links, and is complementary to other information retrieval methods. [1] The development of citation indexing and the launching of citation databases by the Institute for Scientific Information (ISI) during the '60s have been fundamental for the development of citation analytical methods, in particular citation based science mapping (Garfield, 1998).

Citation based science mapping is an area of bibliometrics where the structure and development of science are elaborated and visualized through the analysis of bibliographic data, representing research documents, mostly articles. The objective for citation based science mapping has commonly been to reveal the cognitive structure of science in terms of visualizing and describing its sub-division in disciplines (fields), sub-disciplines and specialties. Also, mappings have been focusing on scientific information provision (e.g. the ISI product Atlas of Science [2]). The notions of discipline, sub-discipline and specialty should be clarified. A discipline should be the broadest entity, denoting a branch of scholarly knowledge, e.g. physics. Physics in turn can be divided in sub-disciplines like condensed matter physics which in turn can

[1] "Information retrieval deals with the representation, storage, organization of and access to information items" (Baeza-Yates & Ribeiro-Neto, 1999, p. 1).

[2] The Atlas of Science was presented in 1981and was based on the clustering of highly cited and cocited documents from a given sub-specialty and provided the user with a mini-review of the subject, a bibliography of clustered documents, a cluster-map depicting the documents in a cluster - the similarity or distance between them - and a bibliography of documents citing the clustered documents.

be divided in specialties like solid state physics, materials physics and polymer physics, which in turn can be divided in other (sub-) specialties. These terms reflect a *function* of continuous specialization, subdivision and new amalgamations of research over time, rather than well demarcated and static hierarchical levels of classification. This function of specialization is due to the fact that a single researcher can not attain a detailed knowledge of all areas within a certain discipline. Hence, by necessity researchers must focus on a specific area within their fields or sub-fields. Those researchers with a common focus communicate (both formally [through academic journals] and informally) and over time such a group with a specialized research focus form an area of specialization. The term "field" is frequently used in the literature and may cover any demarcated area of research.[3]

In citation based science mapping, different entities (journals, authors or documents) in bibliographic descriptions representing research documents are applied as analyzed units for different purposes. For instance, when mapping[4] citation relationships between journals, an overall view of the discipline structure of science may be arrived at. However, the journal is a too broad a unit of analysis to reveal the fine structure of science (Small, 1974). Hence, citation based mapping with the objective to map specialties usually employs documents as the unit of analysis and it has been suggested that the "[s]pecialty is the principal mode of social and cognitive organization in modern science" (Small, 1977).

The usefulness of science mapping is clear as "most scientists have intuitive notions about the subdivisions of their fields, but no observer, however broadly trained, can gain an overall perspective in the scientific mosaic" (Small, 1974). The difficulty for researchers to gain an intellectual key map over their own discipline's subdivision in specialties and research foci within specialties is augmented by the increasingly interdisciplinary character of research where new lines of research transcend boarders between disciplines. A good example of this is the (non-traditional) "field" of environmental science, which connects several disciplines and sub-disciplines like astrophysics, chemistry, ecology etc. Conclusively, the mapping of research specialties may provide means, not only for the study of the specialty structure of science, but also for new approaches of indexing and information provision for scientists (cf. Small, 1973).

Historically, the development of citation based science mapping is associated with experiments that were launched in the '70s by ISI where the mapping method was cocitation cluster analysis. This method is defined by the measure of document similarity and the method of clustering applied. The measure of document similarity is the cocitation of documents and single link clustering is the cluster method. Though several improvements of the cocitation cluster technique have been accomplished over the years, the method of document cocitation clustering has been criticized on

---

[3] The general term "field" mostly denotes the discipline level or the specialty level, depending on the context. It is often difficult to classify the exact level of scientific activity and the use of terms in the literature is ambigious and inconsistent. The terms"sub-field" and "sub-specialties" are sometimes used as well.

[4] It should be noted that the verb "map" indicates that something is mapped, while the noun "map" stands for a graphical representation that may enhance our spatial understanding of associations between objects. Hence, in the context of science mapping, mapping need not lead to maps, though it often does.

methodological grounds (Leydesdorff, 1987; Oberski, 1988). The advocates of this method claim that the fine structure of science in terms of identified and mapped specialties is reflected. This has been seriously questioned on grounds of statistical instability resulting from arbitrary application of threshold settings and the use of the single link cluster method. In spite of the criticism, the basic application of document cocitation clustering has not changed and is still dominating as at today.

On the other hand, there exists another citation based measure of document similarity, namely, bibliographic coupling, which was introduced to the research community in the early '60s (Kessler, 1962 and 1963a). In comparison with the cocitation approach, bibliographic coupling methods have the advantage of being capable of identifying emerging specialties (Glänzel & Czerwon, 1995 and 1996), as research articles are available for analysis as soon as they are published. In the case of cocitation analysis, there will always be a time lag between the current published research and the generation of a sufficient number of received citations that can facilitate stable sets of cocitation data for mapping. However, there is also another distinct difference between the cocitation and bibliographic coupling approaches. With regard to cocitation, claims of the identification and mapping of research specialties is based on the presumption that highly cited documents represent central concepts of specialties and that the grouping of such highly cited items on basis of cocitation therefore would reflect the cognitive structures of specialties. With regard to bibliographic coupling, claims can generally not be made that articles represent central concepts as no immediately applicable criterion for this exists. Hence, applying bibliographic coupling for mapping purposes, one could generally not make the same claim of identifying the cognitive structure of a research specialty. This means that cocitation analysis and bibliographic coupling should be complementary to each other.

Despite its favorable features, there is a distinct lack of evaluative research concerning bibliographic coupling applied as a science mapping method. The reasons for this unobtrusive position in science mapping are not obvious and comparable and complementary results to the cocitation approach have also been reported when this measure was applied for science mapping purposes (Sharabchiev, 1988; Persson, 1994; Jarneving, 2001). In addition, research in bibliographic coupling has shown that the identification of "hot" research areas could be accomplished by the identification of "core documents", i.e. currently published research articles with many and strong associations of bibliographic coupling to other currently published research articles, and that most core documents belong to a few high impact documents of a specialty (Glänzel & Czerwon, 1996).

For citation based science mapping in general, it also holds that only a small fraction of articles of a selected original population is available for mapping as citation based science mapping depends on consensual referencing. This means that a lack of consensus about which previous research that is the most significant in relation to a common topic, or less attentive referencing, would lead to a loss of cognitive association between articles and a diminishing of the original population (cf. Braam, Moed & van Raan, 1991). This concerns the extent of exhaustiveness of mapping results and affects the validity of claims of identification and mapping of specialties.

Conclusively, citation based science mapping is generally attached with uncertainty when the objective is set to identify and define the specialty structure of science. With

regard to information provision or information sharing objectives, this uncertainty should have lesser importance as the currency and relevance of obtained information should be the first priority, not the exactness of the mirroring of specialty cognitive structures.

Based on the findings of the various researches so far, bibliographic coupling could be combined with a cluster method to provide a method of science mapping complementary to the prevailing cocitation cluster analytical method. The complete link cluster method would on theoretical grounds (cf. Everitt, Landau & Leese, p. 60-62) provide a suitable cluster method for this purpose, for more coherent clusters would be generated, meaning that it would not have the drawbacks of the single link cluster method. Thus, based on empirical evidence and theoretical considerations, bibliographic coupling and the complete link cluster method were combined to form a method of science mapping which was then evaluated in this study.

The objective was set to evaluate the proposed method's capability to generate subject coherent clusters, i.e. to identify coherent research themes, and the assumed context of application was scientific information provision. The research design comprised four different research settings of which three correspond to specific fields of research and one to a large multidisciplinary research setting, where the specific objective was to identify and apply core documents for the evaluation of the applicability of the proposed method.

Conclusively, the method to be evaluated has the following two primary components:

i. a measure for the association of documents where the association can be expressed as the similarity between two documents; and

ii. a cluster analytical method for the partition of sets (populations) of documents.

The measure of document similarity is needed for the purpose of establishing cognitive relationships between documents. The cluster method is needed for the partition of a set of documents into subsets of reciprocally similar documents. In this study, *bibliographic coupling* is applied as the measure of document similarity and the *complete link cluster method* is used for the clustering of documents.

The whole research process and its findings are presented in five subsequent chapters beginning with Chapter 2, in which the framework of the thesis is presented. In Chapter 3, the research design, the rationales and the research questions are given. Chapter 4 presents bibliometric and statistical methods applied in this study, the methods of data selection and collection as well as the properties of the data collected. Chapter 5 sets out the findings of the study whilst Chapter 6 discusses the findings and gives the conclusions. In order to facilitate the reading, a list of equations discussed in the thesis is given in Appendix 1.

CHAPTER 2: THE THEORETICAL FRAMEWORK

In this chapter, the framework on which the design of the study is based is accounted for. It begins with an elaboration of some concepts which are central to the study. Next, the previous research on which the study builds is presented where the outline of the development of cocitation analysis and bibliographic coupling is given. The purpose of presenting both methods is foremost due to the claim made in this thesis that the proposed method would be complementary to the cocitation cluster analytical method. Another motive is that little empirical experience exists concerning bibliographic coupling in the context of science mapping, whereas the development of cocitation analysis follows a clearly discernable track with a series of connected articles on science mapping. This means that experience of citation based science mapping on the document level must be derived from empirical findings from cocitation analysis.

The chapter ends with a summary and a discussion of the foundation for the research design of this study.

1. CENTRAL CONCEPTS

1.1 Citation Indexing

Citation indexing was developed as a result of the needs of scientists to find and reference previous published research. A citation index lists documents that have been cited and identifies the sources of the citation. The strength of citation indexing is its simplicity. Just by knowing an item that has been cited, several additional documents can be found. Semantic difficulties are avoided as citation symbols rather than words are used to describe the content of a document. This makes the job of the researcher easier when searching for works from other disciplines, as they are not required to know the terminology of the disciplines that they are searching in order to make the search.

Traditional subject indexing involves specialist judgment, increasing the time and the cost of indexing with increasing indexing depth[5]. Citation indexing solves the depth versus cost problem by substituting the author's citations for the indexer's judgments and there are no restrictions as to the number of citations (the reason why citation indexing in most cases should be deeper than subject indexing where a few indexing terms are used). Also important is that citations are timeliness, whereas the usability of an indexing term is due to semantic stability meaning that the actuality of indexing terms might be low in subject indexes, thus, limiting their effectiveness as search tools (Garfield, 1979, p.1).

In 1961, the database publishing company ISI started to publish the *Science Citation Index* (SCI) and in 1966 it publishes the *Social Science Citation Index* (SSCI). The SCI provides access to 3,700 technical and science journals and the SSCI covers 1,700 social science journals. In 1976, subsequently, ISI

[5] "Indexing depth" aims at the degree to which a topic is represented in detail.

started to publish the *Arts and Humanities Citation Index* (A & HCI), which provides access to 1,130 arts & humanities journals. It should be noted that the ISI databases are multidisciplinary, whereas traditional indexing and abstracting services provide databases that are limited to a single field.

The SCI and the SSCI have consistently been used by the vast majority of research that applies citation based mapping techniques. The A & HCI has also been used but to a considerably lesser extent. Citation data is made accessible either by downloading hundreds or thousands of bibliographic records from citation databases, or through online techniques (cf. Persson, 1988). In this study, data from the SCI and the SSCI are used.

1.2 Citation Analysis

Citation analysis is the area of bibliometrics which deals with the study of the relationships between items of the scientific literature. Several areas of the successful applications of citation analysis have been developed. They include science mapping, information retrieval (IR), evaluation of scientific activity, collection management and history of science. Below is a brief description of these areas of application of citation analysis:

- **Science mapping**

  This concerns the mapping of literature on different levels of scale. Commonly, the structure of particular science fields (specialties) are mapped and elaborated graphical depictions of the relations between important nodes (documents, authors, journals or other types of entities) in the citation network are analyzed. Sometimes, the mapping involves the characteristics of a certain field's literature, and may concern, for example, distribution of citations over language areas, geographical areas and subject areas. Science mapping could also involve the association between disciplines and research fields as well as the development of a science field over time. Science mapping is useful to information professionals involved in the organization of scientific information and it is also an important tool for the monitoring of scientific development.

- **Information Retrieval**

  Citations are considered as useful supplements to keywords in the retrieval of relevant documents and have been used in various retrieval algorithms as well as in the development of document representations. Also, citation analytical methods have been applied to visualize overviews of document collections and have been implemented in Web-based applications.

- **Evaluation of scientific activity**

  Here, citation counts are used as indicators of influence on research and citation analysis is applied as an evaluative tool by science

administrators for the assessment of universities, countries and other aggregates of scientific activity.

- **Collection Management**

  Citation analysis has mainly been applied for the development of journal collections in libraries. Decisions regarding the acquisition, discontinuation or continuation of journals are supported by citation data.

- **History of Science**

  Historical events of scientific enterprise could be traced chronologically by citation relations between central works and the relationship between discoveries is established through the linking of key documents through time.

However, citation analysis has its limitations, which include the assumptions that have to be made in the analysis and also problems associated with citation data and sources, as discussed below.

1.2.1 Basic Assumptions Underlying Citation Analysis

It is difficult to establish the underlying motivations and the significance for a citation, and they can probably never be fully elucidated. As such, one has to rely on some general assumptions. In Smith (1981, p.86 ff.), several assumptions concerning the significance and function of citing are elaborated, of which four of the more pertinent issues are quoted and discussed here.

i. Citation of a document implies use of that document by the citing author

This assumption incorporates that the author refers to the major part of documents used in the preparation of the citing work and that all referenced items were used. Whether a certain item is just quoted without further reading or to what extent the cited item is used, is hard if possible at all to decide.

ii. Citation of a document (author, journal, etc.) reflects the merit (quality, significance, impact) of that document (author, journal, etc.)

The underlying assumption in the use of citation counts as quality indicators is that there is a high positive correlation between the number of citations received and the quality. Arguments concerning the invalidity of citation counts as indicators of quality focus on the fact that documents can be cited for reasons irrelevant to their merit (e.g. negative citations). However, several studies have shown support for citation counts as quality indicators. The operationalization of other measures (non-bibliometric) of quality in comparison is found to be

problematic and Smith (ibid.) concludes that citation counts is are rough measures of quality. Also, one could have more confidence in counts of larger units than on individual counts. Cole & Cole (1973, p. 35 f.) also argued in favor of citation counts as indicators of quality. They reported that "[d]ata available indicate that straight citation counts are highly correlated with virtually every refined measure of quality". They also warned about the misuse of citation counts, i.e. to interpret small differences as significant, and conclude that "[c]itation counts should not be used as fine measures of quality" as small differences should not be interpreted as significant.

iii. Citations are made to the best possible works

A better expression is perhaps the citation of "the most relevant works" in relation to the topic treated by the citing author. However, this assumption may sometimes be wrong as it has been shown that accessibility may be an important factor in the selection of references (Soper, See Smith, 1981) meaning that what is found may not always be the most relevant item. Accessibility, according to Smith, may be a function of form, place of origin, age and language and "it may be that anything that enhances the researcher's visibility is likely to increase his citation rate..." (1981).

iv. All citations are equal

Taken as a major premise is that there is a cognitive relationship between the citing and the cited document. However, the strength of the cognitive relations between the citing and the cited document should not all be the same. The exact nature and strength of such a relationship is hard to characterize and measure. In spite of this, all references of a document are commonly considered to have the same status when used in citation analysis.

Note though that the assumptions are not of equal importance to the different types of citation studies and this needs to be further elaborated. With regard to (i), the use of (the major part) a document is basic both for cocitation analysis and bibliographic coupling as a cosmetic referencing may not reflect the cognitive association between the citing and the cited document, bibliographically coupled documents or between the cocited documents in a valid way.

Point (ii) should be essential for cocitation analysis as high citation counts of cited documents are considered to identify documents as concept markers and are applied as a prime selection criterion for cited documents to be included in the analyses. With regard to analyses of bibliographically coupled articles, point (ii) is of lesser significance as primarily the similarity between reference lists of two coupled articles are considered, not the citation impact of references.

Point (iii) is relevant to both bibliographic coupling and cocitation analysis. This is so, as less attentive or random referencing may lead to the absence of identified cognitive associations between citing documents treating the same topic in the case of bibliographic coupling (cf. Braam, Moed & van Raan, 1991) and in the case of cocitation analysis, less relevant associations between cocited documents would arise.

Lastly, point (iv) points to a problem that should be common to both cocitation analysis and bibliographic coupling. As for now, no practicable method exists for discerning the more important associations between cocited works in a reference list, neither is there a method for the decision of which references common to bibliographically coupled articles that are the more important ones (cf. Martyn, 1964).

1.2.2 Problems of Citation Data and Sources

Objections against the use of citation data in different kinds of studies might have their point of departure in the violation of assumptions used, but there also exist objections that concern the sources themselves, both with respect to citation data and to the ISI citation indexes. With reference to Smith (1981) and Vinkler (1986), thirteen problems concerning sources are mentioned in Egghe & Rousseau (1990, p. 217 ff.). Those problems that are of importance to the application of the proposed method are quoted and commented here.

i. Errors

This refers to errors such as misspelling, incorrect page numbers etc., due to author mistakes and transcription errors. "Whether such problems would cause appreciable error is not known, but probably they would not since there is no reason to suspect that they are systematic" (MacRoberts & MacRoberts, 1989). Systematic errors, on the other hand, could cause problems such as underestimation of citations, for example, preprints can only be indexed under "in press" or "unpublished".

ii. Synonyms

This problem is foremost associated with the way the author's name is being cited. The problem may arise under the following circumstances:

- authors have the same surname but different initials;
- a woman author may be cited in her maiden and married names;
- different transliterations of non Anglo-Saxon names; and
- misspellings.

Also, variations of the abbreviated title of journal names in the reference lists of bibliographic descriptions of the citation indexes are common.

iii. The incompleteness of the ISI databases

As the ISI method of obtaining comprehensive coverage of the literature is based on Bradford's law, which states that only a small percentage of journals account for a large percentage of the significant articles in any given field of science, a consequence is that most journals and articles are not included. Though the body of important research in any field might be well covered, the ISI data might not fulfill the needs of local studies.

iv. The dominance of English as a scientific language

It is clearly so that the English language dominates the scientific communication in the Western world. A consequence is that scientific articles published in English are preferred for citations.

v. The American bias

The citation indexes are known to be biased towards publications from the USA.

With regard to points (i) and (ii), technically, the whole text string identifying a cited reference in a bibliographic record is compared with every other such string in all other bibliographic records representing a population under study. Hence, when two text strings refer to same reference but are not completely identical, such a unit of bibliographical coupling will be omitted, if not standardized to one form. With regard to relatively small populations of source articles, semi-automatic routines may be applied for standardization purposes, increasing the number of bibliographic coupling units (Persson, 1994).

Points (iii) to (v) are of no immediate importance for the evaluation of the proposed method. However, when comprehensive and exhaustive mappings are aimed at, claims of coverage of a field of research may be less valid if a considerable amount of published research is omitted on grounds of incomplete coverage, geographical or language biases.

1.2.3 Citation Based Science Mapping

The data on which mapping and the generation of maps are based on is commonly derived from bibliographical citation databases where research articles are indexed and made accessible as bibliographic records. A *bibliographic record* is a representation of a research article, and contains less information than the item it represents. The information contained in a bibliographic record usually tells us who authored the article, where and when it was published as well as its subject content as indicated by abstract, title, journal title, classification codes, author key-words and assigned descriptor terms. [6] The type of bibliographic records used in this study not only provides the aforesaid information but also contain *references* which link to the previous research that is referred to in research articles. A reference is given to a work cited in an article and is counted only once, as it occurs in the reference list of the article. One way to distinguish between references and *citations* is that references in a document is a property of the same, while the citation of a document informs us about the extent to which it is noticed by subsequent researchers. This is of some importance as one sometimes maps the cited works and sometimes the citing works.

In document based bibliometric mapping, citation based measures of the association between documents are applied. There are three forms of citation associations between documents as follows:

i. direct citations;

ii. cocitations; and

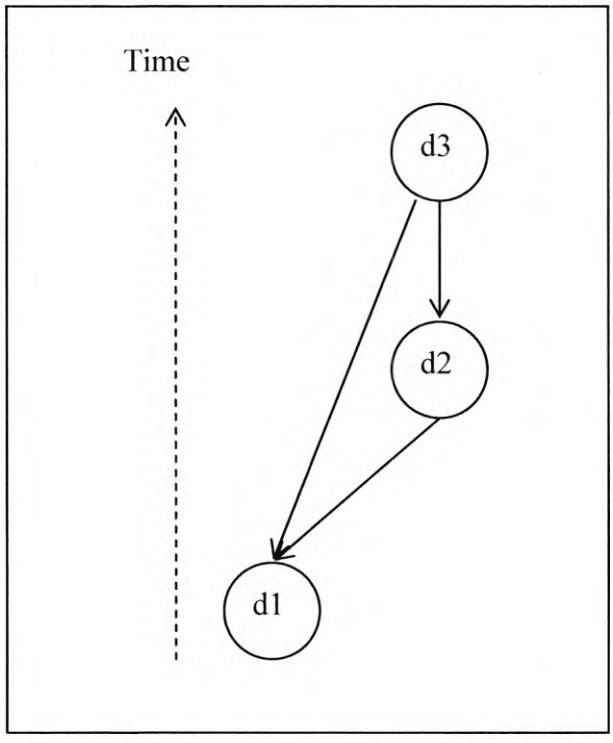
iii. bibliographic couplings.

Direct citations means that a document is cited in another document and the strength of the association between two documents is either 0 or 1. An association of cocitation between two documents means that both documents are cited together in other documents, hence, the association is generated extrinsic to the associated documents. The strength of association between a pair of cocited documents is 1…n, depending on the number of times they have been cited together. A bibliographic coupling between two documents means that both documents cite the same third document. The association between two bibliographically coupled documents is intrinsic to the documents and the strength of association is 1…n, depending on the number of common references. Generally, the association (coupling) between two

[6] Several terms denoting scientific, published works are incorporated in the bibliometric jargon, namely, article, document and publication. When referring to original texts, their authors' choices of terms will be applied. The term "document" covers for other document types besides journal articles, and is applied when motivated, otherwise, the term "article" is applied. It is to be noted that though the citation databases of ISI only index journal articles (the citing items), the articles contain references (the cited items) directed to any document type. Though bibliographic descriptions of journal articles (bibliographic records) are used as input data in computational operations and calculations, rather than articles, conceptually, journal articles are analysed and are referred to also when bibliographic records are treated in practice.

documents is referred to as a *link*. A graphical illustration of the three types of citation associations are given in Figure 2-1.

Figure 2-1: The Citation Associations Between Three Documents



The three documents in Figure 2-1, i.e., d1, d2 and d3, are published at different points in time. All three documents are associated through direct citations. Two types of document pairs are formed from them. The first pair (d1 – d2) is generated through citations from d3 (cocitation). The second pair (d2 – d3) is generated through their common referencing of d1 (bibliographic coupling).

As the vocabulary of bibliometric mapping research is partly confusing, the separation between the concepts of measure and method are seldom clearly reflected by the use of the terms. The terms *bibliographic coupling* and *cocitation* denote *measures* of document association. When applying these measures, one arrives at values of *bibliographic coupling strength* and *cocitation frequency*. In the literature, the term *cocitation analysis* usually denotes method applications where cocitation relations are analyzed, mostly for science mapping purposes.

The strength of association generated by either bibliographic coupling or cocitation is to be considered as the perceived *similarity* or *distance* between two documents where the strength of similarity is inversely related to the distance, i.e. a short distance corresponds to a high similarity and vice versa. A variety of statistical mapping techniques can be applied where input data is the values of cocitation frequency or bibliographic coupling strength, or normalized values of the same. The result is commonly a categorization of

documents where documents sharing a common research focus are gathered in *clusters*.

The general definition of a cluster is a *group* of objects. However, in this study, the term "cluster" mostly refers to the *partition* of a set of research articles into subsets by means of some cluster analytical method (see Sub-section 1.4 in this chapter). The size of a subset can vary between 1 and n and a subset containing only one element is named *singleton cluster*. Also, the concept of *cluster relevance* needs some clarification. Generally, relevance is about how pertinent or connected certain information is to a given matter. When the relevance of a cluster is assessed, this concerns how well the cluster represents a coherent research theme, and different variables are applied for the measurement and assessment of relevance (see Sub-section 2.2 in Chapter 4).

Other methods than cluster analytical (applying the same kind of data) may project cognitive associations between objects in a two or three-dimensional display, so that the distance between points in the projection represents the similarity between the objects. Such a method is called multidimensional scaling (MDS). A more detailed elaboration of MDS is given under Sub-section 2.5 in Chapter 4.

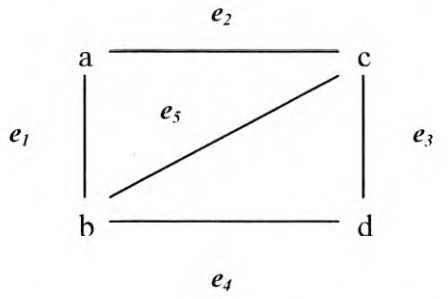
1.3 Mathematical Concepts and Definitions

The understanding of citation associations may be enhanced by applying concepts that are applicable to networks in general. Graph theory supplies such concepts. As such, in this study, different sets of bibliographically coupled documents (e.g. clusters) will be considered as networks which may be depicted as *graphs*.

An *undirected graph G,* is constituted by a set V of *vertices* and a set E of *edges* such that each edge $e \in E$ is associated with an unordered pair of vertices.[7] The existence of an unique edge e associated with the vertices v and w, implies the existence of an edge e associated with the vertices w and v and this is written as $e = (v, w)$ or $e = (w, v)$ (Johnsonbaugh, 1997, p. 306). In Figure 2-2, is an example of an undirected graph G. It consists of the set V = $\{a, b, c, d\}$ of vertices and the set E = $\{e_1, e_2 \ldots, e_5\}$ of edges.

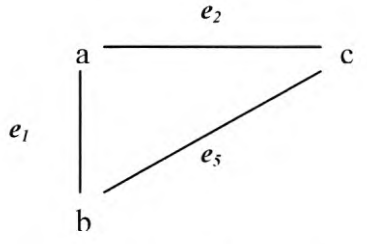
[7] The terms used in relation to graphs, namely, "vertice", and "edge" correspond to documents and the bibliographic coupling between two documents respectively. In more general discussions concerning clusters and their associations through bibliographic coupling, the corresponding terms "articles" and "links" are used.

Figure 2-2: The Undirected Graph G



A graph G' whose vertices and edges form subsets of the vertices and graph edges of a graph G, is a *subgraph* of G, and G is said to be a *supergraph* of G'. A *complete graph* is a graph in which each pair of vertices is connected by an edge. In Figure 2-3, subsets of V and E constitute the subgraph G', which also is a complete graph.

Figure 2-3: The Subgraph G′ of the Undirected Graph G



An undirected graph can be presented by a *symmetrical matrix*. A matrix M, is a rectangular array of numbers, where M has m rows and n columns and the size of M is m x n. The numbers pertain to the elements of V and they are represented by the letters i and j and it is assumed that i and j run from 1 to n. The number connecting i with j is represented by m_{ij}.

A *square* matrix is one where the number of rows and columns are equal, n x n, and a symmetrical matrix is a square matrix where $m_{ij} = m_{ji}$. Hence, the associations between the elements of V can be represented. The columns and rows are labeled with the elements in V and m_{ij} is equal to 1 if there is an edge between the vertices of the elements in V and 0 when there is no edge between the vertices of the elements in V (see Table 2-1).

Table 2-1: The Undirected Graph *G* Represented by a Symmetrical Matrix

	a	**b**	**c**	**d**
a	0	1	1	0
b	1	0	1	1
c	1	1	0	1
d	0	1	1	0

Note: The diagonal elements indicate the associations between i and i which are of no importance in this case. Only half the matrix is needed, (below or above the diagonal) as $m_{ij} = m_{ji}$.

When analyzing graphs and matrices, it is necessary to know some counting methods. The first is the *multiplication principle* which states that if an activity can be constructed in t successive steps and step 1 can be done in n_1 ways; step 2 in n_2 ways and step t in n_t ways, then the number of different ways is $n_1 \cdot n_2 \cdots n_t$.

The second principle is *permutation,* which is related to the order of objects. In concordance with the principle of multiplication, the first object can (for example) be selected in four ways, the second in n - 1 ways, the third in n - 2 ways and so on. Hence, there are $n(n - 1)(n - 2)\cdots 2 \cdot 1 = n!$ permutations of n objects (ibid. p. 210).

An r *permutation* of n distinct elements $x_1 \ldots x_n$ is an ordering of an r- element subset of $\{x_1 \ldots x_n\}$. The number of r-permutations of a set of distinct elements is denoted by $P(n, r)$ and $P(n, r) = n(n - 1)(n - 2) \cdots (n - r + 1)$. When one selects objects without regard to order, it is a *combination.* An r *combination* of n distinct elements $x_1 \ldots x_n$ is an unordered selection of an r-element subset of $\{x_1 \ldots x_n\}$. The number of r-combinations of a set of n distinct elements is denoted by $C(n, r)$ or $\binom{n}{r}$

and

$$C(n,r) = \frac{P(n,r)}{r!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{(n-r)!\,r!}$$ (ibid. p. 211-213). **(2.1)**

1.4 Classification and Cluster Analysis

The second component of the two constituting the proposed method for science mapping is the method of partition. The idea of mapping science on the basis of published research articles implies a method of partition where objects are grouped to produce a classification. A classification should then fulfill the following conditions:

i. it should be exhaustive; and

ii. classes should be mutually exclusive.

This means that each object should belong to exactly one class. The forming of classes should also imply that classified objects are more similar to other objects in the same class than to objects in another class. The objective of finding such classes connects with the purpose of a set of statistical techniques with the generic name "cluster analysis". Hence, cluster analysis involves techniques that produce classifications from data that are initially unclassified. From another point of view, cluster analysis is essentially about discovering groups in data (Everitt, Landau & Leese, 2001, p. 6).

Cluster analysis is highly empirical and different methods can lead to different groupings, both in number and in content. This happens because the choice of cluster algorithm imposes a structure and cluster methods might detect clusters that have no correspondence to the real world. It is usually difficult to judge if the results make sense in the context of the problem being studied (ibid.). This concerns the fact that there are many cluster algorithms but no generally accepted best method and there is usually a subjective component in the assessment of the results. The task is, therefore, to select the most appropriate method in relation to data and empirical experiences.

1.4.1 Cluster Analytical Methods

The commonly used methods fall into the following two general categories:

i. non-hierarchical; and

ii. hierarchical.

The *non-hierarchical* approach requires that some objects be selected as cluster seed points around which clusters are then built. This is accomplished by assigning every object in the population to its closest cluster seed object. After this step, clusters may be split, and clusters close to one another may be combined. That is, objects are allowed to move in and out of groups at different stages of the analysis. This approach has some disadvantages according to Johnson (1998, p. 323). They include:

i. it requires one to initially guess the number of clusters that is going to exist;

ii. it is greatly influenced by the choice of the initial cluster seed objects. By letting the statistical program choose the seeds, the selection often depends on the order in which the data are read into the computer. As such, two researchers could perform a cluster analysis on the same set of data and produce entirely different clusters; and

iii. the procedure is often not feasible computationally because there are just too many possible choices in terms of number of clusters and number of locations of the clusters seeds.

In bibliometric mapping, the numbers of clusters are usually not known beforehand, which makes non-hierarchical cluster methods less applicable.

In general, the most widely used cluster methods are the *hierarchical* ones. In hierarchical methods, groups are formed by a process of agglomeration or division. The agglomeration process starts with all objects being alone in groups of one, that is, each object is considered a cluster (a singleton cluster). Objects are then gradually merged according to some algorithm until finally all individuals are in one group. The process of division begins with all objects being in one group. This is then split into two groups; the two groups are then split, and so on until all objects are in groups of their own.

The general procedure of hierarchical agglomerative methods starts with the compilation of a matrix of proximity values showing similarity or dissimilarity. For example, let M be an N·N squared proximity matrix and let N clusters contain one object each and the clusters denoted 1 to N. Next, apply a scheme of agglomeration where all objects begin alone in groups of size one and groups that are "close" (similar) together are fused according to the steps presented below: [8]

i. Find the most similar pair of clusters i and j. Denote this similarity M_{ij}.

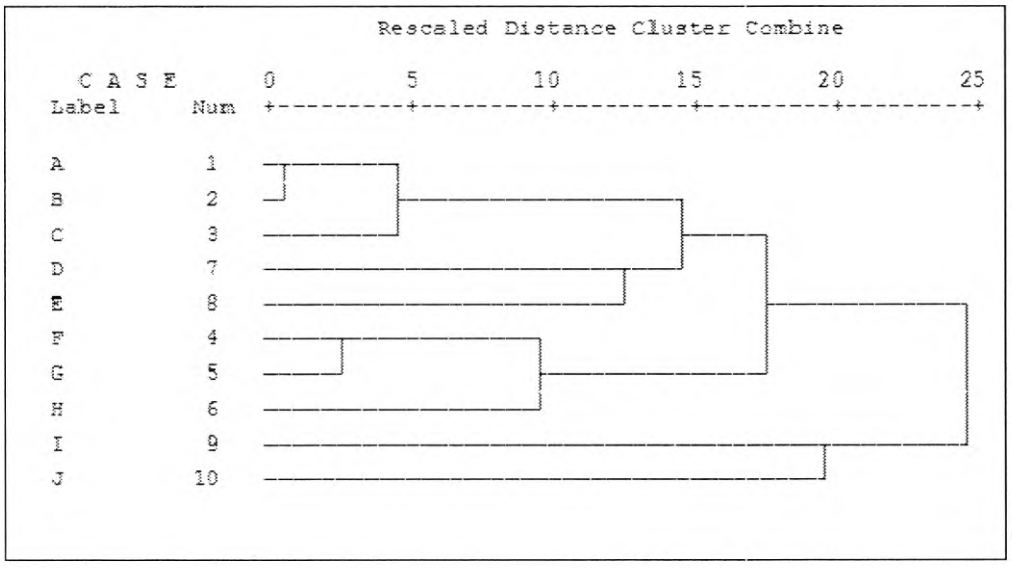
ii. Reduce the number of clusters by one through the fusion of clusters i and j. Name the new cluster p ($= j$) and update the matrix according to the revised proximity between cluster p and all other clusters.

iii. Repeat steps (i) and (ii) until all objects are in one cluster.

The result of the cluster process can be visualized by a dendrogram. A dendrogram is a two-dimensional tree-diagram which illustrates the fusions of clusters at different levels of distance at each stage of the analysis. The nodes in the dendrogram (the point where two lines meet) represent clusters and similar clusters are joined by links whose position in the diagram is determined by the level of similarity between them. An example of a dendrogram is given in Figure 2-4.

[8] Adapted from SPSS technical papers: Clustering Methods/ general procedure)

Figure 2-4: Example of a Dendrogram



Note: This is a fusion of 10 objects, A to J. The dendrogram rescales the actual distances to numbers between 0 and 25, preserving the ratio of the distances between steps. The closest objects are A and B, which are merged with C in the next step, etc.

When selecting an appropriate method of clustering, experience achieved and recorded by researchers within the field could to some extent be used as a guide. Originally, when the cocitation clustering method was developed by Henry Small and colleagues at the ISI (Small,1973; Small & Griffith, 1974; Griffith, Small, Stonehill & Dey, 1974; Small & Sweeney, 1985), the *single link* cluster method was applied. The defining feature of this method is that the distance between groups is defined as that of the closest pair of individuals. Single link clustering, is known to produce straggling and loosely bound clusters, especially in large data sets, and this problem might show up as a less clear structure due to this "chaining" phenomenon. Still, single link applications seem to have been successfully used by many researchers in the context of document cocitation analysis (e.g. Small & Griffith, 1974; Griffith, Small, Stonehill & Dey, 1974; Small & Griffith, 1983; Small & Sweeney, 1985; Braam, Moed & van Raan, 1991) and a variant was used by Persson performing an author cocitation analysis[9] (1994). The single link method is easy to implement and use especially when large amounts of data is to be clustered. However, as the development of cocitation cluster analysis method has attracted criticism (Leydesdorff, 1987), the use of *Ward's* method has been suggested as an alternative. Ward's method has also been mentioned as appropriate in the context of author cocitation analysis (McCain, 1990) as has the *complete link* cluster method (McCain, 1990; White & McCain, 1998). Ward's method differs radically from the single link and complete link cluster methods as distances between objects are defined rather than differences between clusters and at each stage of the clustering process, the objective is to minimize the increase in the total within–cluster error sum of squares. This

[9] Author cocitation analysis is a special case of cocitation analysis where the analyzed units are the authors' names in referenced works. In author cocitation analysis, the collected research of an author is represented by the author's name.

method is usually used with a distance matrix of proximity data and a matrix of squared Euclidean distances is required as this method assumes that objects can be represented in Euclidean space (Everitt, Landau & Leese, 2001. p. 62).[10]

Comparing complete link clustering with single link clustering, the difference is how the distance between an existing cluster and a candidate object for fusion with that cluster is defined. In complete link clustering, the largest distance between the candidate object and any object of the existing cluster is sought. This means that any candidate must be within a certain level of similarity to *all* members of that cluster. As mentioned, in single link clustering the shortest distance between clusters is sought. Hence, single link clustering and complete link clustering could be seen as each other's opposites. In addition to these methods, the *between groups average link* appears as an alternative in this study. For this method, the distance between two clusters is the average of the distance between all pairs of individuals that are made up by one individual from each group. It was developed as an "antidote" to the extremes of both single and complete link (Aldenderfer & Blashfield, 1984, p. 40). In theory, it is possible to make some general assumptions concerning clusters generated by these methods according to differences between their algorithms. Hence, the single link cluster method would generate more loosely bound clusters whereas the complete link cluster method would produce compact clusters and the group average link method something in-between.

1.4.2 Motives for the Choice of the Complete Link Cluster Method

If one can assume that the similarity between document A and document B and the similarity between document B and document C generally implies a similarity between document A and document C, a method of clustering with less severe conditions to fulfill may be appropriate (e.g. the single link method). This assumption, however, should be considered unconfirmed as previous research in cocitation clustering has shown that the chaining effect of the single link method has caused some undesirable effects like large subject inconsistent clusters when applied for cocitation cluster analysis (see Sub-section 2.2 in this chapter). Though cocitation and bibliographic coupling are not the same, they are similar in the sense that they are both based on consensual referencing. Thus, it does not seem too far-fetched to assume that similar drawbacks might occur when the single link method is applied on bibliographic coupling data. It should also be noted that reciprocal associations between documents in a cluster through bibliographic coupling do not necessarily imply that one single cited reference is common to all the documents. Moreover, as discussed, the significance of the association between two documents through a common reference (a bibliographic coupling unit) is hard to establish. All this speaks in favor of not introducing further uncertainty on an additional level. Therefore, a method that ensures that all objects in a cluster are within a set maximal distance to each other would be preferred in order to secure coherent clusters.

[10] The need for using Euclidean distances motivated the exclusion of this cluster method as a candidate method for partition. See Sub-section 1.1 in Chapter 4 concerning proximity measures.

From a graph theoretical viewpoint, such groups could be considered complete undirected graphs. Such graphs would always have a maximal degree of interconnectedness, i.e. a maximal *density* (*D*), where *D* is defined as:

$$D = \frac{2 \cdot (\# L(G)}{N(N-1)}, \qquad \textbf{(2.2)}$$

where

$\#L(G)$= the number of edges connecting two vertices; and

N= the number of vertices (Otte & Rousseau, 2002).

The interval is [0, 1] and the maximum value is reached when the value of $\#L(G)$ equals the value of $N(N\text{-}1)/2$. In this context this means that the maximum value is reached when all possible document pairs in a cluster are bibliographically coupled.

Applying the complete link cluster method, one will arrive at clusters with a maximal density *D*, as each cluster member is associated with every other member in a cluster, given that fusions of clusters at a level of zero association are prohibited. [11] As the maximal interconnectedness is given by the method, only the strength of association (the distance) between documents varies. A maximal allowed distance between documents in clusters could be set as a way of avoiding more random associations between articles, and secure a high degree of similarity between articles in clusters.

It was hence presumed that the complete link cluster method would generate coherent clusters and, therefore, also more subject consistent (relevant) clusters. However, one could always argue that this cluster algorithm may lead to a low interception of documents as one could imagine something like nearly complete graphs.

[11] Applying the "furthest neighbour" cluster algorithm in SPSS 11.5, the fusion process continues until all objects belong to one cluster.

2. PREVIOUS RESEARCH

2.1 Bibliographic Coupling

Bibliographic coupling was introduced by Kessler to the scientific society through a number of reports and research articles in the '60s. [12]

Bibliographic coupling was primarily described as a method for grouping technical and scientific documents, facilitating scientific information provision and document retrieval. In one of the early reports, a general outline of the context in which an indexing method, concerned with countable indicators based on references, might operate was given (1960). In a subsequent report, the definition of bibliographic coupling was stated: "[a] single item of reference shared by two documents is defined as a unit of coupling between them" (1962). Based on this unit, two graded criteria of coupling were defined (ibid.):

> Criterion A – A number of articles constitute a related group G_A if each member of the group has at least one reference (one coupling unit) in common with a given test article, P_o. The coupling strength between P_o and any member of G_A is measured by the number of coupling units between them. G^n_A is that portion of G_A that is linked to P_o through n coupling units. (According to this criterion, there need not be any coupling between the members of G_A, only between them and P_o)
>
> Criterion B – A number of articles constitute a related group G_B if each member of the group has at least one coupling unit with every other member of the group. The coupling strength of G_B is measured by the number of coupling units between its members. Criterion B differs from criterion A in that it forms a closed structure of interrelated articles, whereas criterion A forms an open structure of articles related to a test article.

The problems concerning scientific and technical information processes related to the invention of bibliographic coupling emanated from the accelerating growth of scientific and technical activities all over the world. Elaborating on the units of the scientific message, Kessler addressed the problem and need of a refined bunching process that would generate a more differentiated and individualized set of articles which would fit the need of individual researchers and groups (1960). From his point of view, the problem of indexing and the subsequent matching of print material with the scientist's need, was the main problem. In the context of a proposed science communication system, he stated that "[t]he goal is to discover certain measurable or countable indicators that reflect on the operational background of a scientist in terms of the four components previously mentioned" (ibid.). The four components that he referred to were the *Man*, the *System*, the *Operation* and the *Results*, each referring to the formal pattern common to the scientific work. According to him, these four components should be used to index documents. He also suggested that a scientific document is a reflection

[12] The suggestion to group scientific documents on the basis of use rather than content was suggested independently by Fano (1956) and by Kessler (1958).

of the operational history prior to publication and that a scientist's information needs are also determined by his operational background. Therefore, the best way to determine a man's need for information is to examine him in terms of the same four components that were used to index the document. According to him, the component *Man* would be sufficiently described by the cultural environment that best describes the scientist's position in the intellectual community and the reflection of this environment may be found in the bibliography of citations that he finds necessary to append to his remarks. In other words, Kessler suggested that references could serve as countable indicators of the intellectual environment in which scientists find information useful for their current work. One could then assume that a shared intellectual environment (as reflected by common references) between two documents might indicate a relationship and this information might be used to facilitate information provision.

In a subsequent report dated 1962, Kessler applied bibliographic coupling to a test population of 40 documents from the field of radio engineering in order to test if a number of scientific documents bear meaningful relations to one another. He found that bibliographic coupling was able to partition this population into valid, related sub-groups. In order to be able to make any generalizations at all as to larger populations and other fields, he subsequently carried out an experiment where the automatic processing of a population of scientific documents (36 volumes of *Physical Review*) resulted in a grouping of 8,521 articles in concordance with criterion A (1963a).

This experiment had its starting point in the population of 8,521 articles, where common references between each member of the population were sought. Kessler (ibid.) reported the outcome from processing one volume (265 articles) and claimed that this experiment proved the existence of groups of documents $G_A(P_o)$ related to one another through the coupling with a fixed document P_o in a well-established field of science. This process could then be iterated. The outline of application areas of this method concerned indexing, classification and information retrieval. He intellectually and by reference connects this document with the previous report from 1960 by pointing out the significance of this method to information retrieval, i.e., once a particular document is identified as relevant, a retrieval system could also retrieve $G_A(P_o)$. He also highlighted some properties of the method:

- The method is independent of words and language as all the processing is done in terms of numbers.
- No expert judgment is required – the text is in fact abundant.
- The group of documents associated with a fixed document, $G_A(P_o)$, extends into the past as well as the future.
- The size and growth of a G_A will reflect the continuous impact of a fixed document and the groupings will undergo changes that reflect current usage and interests of the scientific community.

- Documents similar in the way that they all share references with a third document could be seen as this document's "logical references", which could be substituted for its real references.

Conclusively, Kessler showed in this paper the existence of subject relatedness between bibliographically coupled documents.

In the same year, Kessler published another paper (1963b) where he further elaborated the application of bibliographic coupling in the context of information retrieval by trying to establish a factual background that could guide the design of an experimental science communication system. Bibliographic coupling was applied to a population of 8,186 articles from the *Physical Review* and reported as ten case histories, each illustrating an information retrieval problem. Different strategies of bibliographic coupling were applied, where the effects of enlarging or diminishing the search span by assigning P_0s serially first (in the case of enlarging) or last (in the case of diminishing) in the list of the available literature were tested. Kessler concluded that bibliographic coupling can be applied to a large body of literature and that the process operates both in the future and in the past, relative to the position of P_0. This showed that bibliographic coupling could be used to identify the life span of a given literature.

In 1965, Kessler, still using data from the *Physical Review* for his experiments, compared groups formed according to the *Analytic Subject Index* and by bibliographic coupling. The aim of this experiment was to investigate how bibliographic coupling compare with results obtained by standard methods. He concluded that there was a high correlation between groups formed by bibliographic coupling and groups formed by analytical subject indexing. However, he pointed out that the report did not pass judgment on the utility of either method to any specific application.

In a review article, Weinberg (1974) covered the major part written on bibliographic coupling up to the publication of her article. She concluded that "[a]t this point, bibliographic coupling does seem to be a useful tool for studying the 'science of science' - citation patterns, the useful life of literature, most cited journals etc". However, she reflected on how citation behavior affects the standardization of the citation "unit" and put forward the meaning that the notion of "meaningful groups", claimed by those who advocates bibliographic coupling, may well constitute a problem. What she was trying to say was that since Kessler's experiments were done on documents in one field, there was already a meaningful group to begin with. Therefore, only a test on the scale of SCI would show if bibliographic coupling would work well in a complex and interdisciplinary environment.

However, the first attempt to test the validity and effectiveness of the bibliographic coupling technique for detecting subject relatedness between documents on a more heterogeneous population of documents and on a large scale, was not performed until more than twenty years after Kessler's 1963 reports. One reason for this was probably the technical restrictions imposed by

existing (at that time) computational resources and the problems in accessing large amounts of citation data.

In 1984, Vladutz and Cook carried out an experiment with 10,000 randomly selected documents from the SCI which served as test documents for which bibliographically coupled publications from the entire 1981 database were sought. The large data file covering a multitude of scientific disciplines used in this experiment, corresponded well with Weinberg's claim in 1974 of an interdisciplinary environment as a prerequisite for the evaluation of bibliographic coupling. The questions to be answered with respect to this experiment concerned the frequency of bibliographic coupling links within the file and the degree to which these links are meaningful. It was found that 90 percent of the input articles that have references, yielded a group of at least two coupled items.

Looking back at Kessler's experiments in the '60s, Vladutz and Cook wanted to test more extensively the hypothesis that strong bibliographic coupling links imply strong subject relatedness. The evaluation of subject relatedness was performed by small groups of experts with a scientific background and trained in assigning brief subject descriptions to groups of documents generated by cocitation clustering. Lists of 300 randomly selected test documents together with their strongest coupled articles were presented to the experts. It was found that in over 85 percent of the cases, the articles proved to be closely related by subject to the test documents. Vladutz and Cook concluded that the utilization of bibliographic coupling in a very large citation database was practically feasible and that valid results as to subject relatedness were achieved. The hypothesis stated in this research was that bibliographical coupling "[m]ay prove to be the easiest approximation to an algorithm for revealing the semantically closest neighbors of publications".

A year before Vladutz and Cook published their results, Sen and Gan (1983) had published a purely theoretical article on bibliographic coupling. Their point of departure was a statement by Martyn in 1964 where he argues "[t]hat bibliographic coupling is not a unit but merely an indication of the existence of the probability, value unknown, of relationship between two documents".[13] The two researchers felt that in spite of the attention that previous works on bibliographic coupling had attracted, the method had hardly been taken seriously and that there was a need for a theoretical elaboration. With a point of departure in an M x N hypothetical Boolean matrix, where elements indicated a citation relationship between rows (citing documents) and columns (cited documents), the grouping of coupled documents in bibliographic *cliques* and *clusters* was elaborated. The notion "*clique*" is here equivalent to Kessler's grouping principle G_B, and "clusters would be formed by the populations which have at least one member having coupling with another member whereas no member of one cluster will have coupling with any member of another separate cluster".

[13] The meaning of this statement in short is that the fact that two documents have a reference in common is no guarantee that both documents are referring to the same piece of information in the cited document. Hence, bibliographic coupling is only an indication of the existence of the probability of relationship between two documents.

With regard to the central issue of cognitive resemblance between bibliographically coupled documents, a measure of coupling strength, the Coupling Angle (C.A.) was suggested. The Coupling Angle was expressed as:

$$C.A. = \frac{(D_{oj} \bullet D_{ok})}{\sqrt{(D_{oj} \cdot D_{oj})(D_{ok} \cdot D_{ok})}} \qquad \textbf{(2.3)}$$

C.A. is the coupling angle for citing documents *j* and *k*. D_{oj} and D_{ok} are the binary vectors of document *j* respectively *k*.

The coupling angle C.A. is a geometric interpretation where the C.A. takes the maximum value of 1 if two Boolean vectors are parallel and 0 if they are rectangular. Two documents may be considered to be concerned with a related topic if the angle between vectors representing documents does not exceed a given angle θ ($0° \leq \theta < 90°$) (Glänzel & Czerwon, 1996). Lacking a theoretical basis as well as empirical evidence for the determination of a threshold of coupling strength, Sen and Gan suggested a semi-arbitrary approach with a cut off value of 0.5, which corresponds to $\theta = 60°$.

The question of cognitive resemblance related to bibliographic coupling was also pursued in Peters, Braam and van Raan (1995). These researchers tried to find out whether relatively strong cognitive resemblance within groups of documents, bibliographically coupled by one and the same highly cited item, is present in an interdisciplinary field, i.e., Chemical Engineering. This was operationalized by measuring word-profile similarities between the citing documents. It was found that word profile similarity within groups sharing a citation to a highly cited publication was significantly higher than between documents without such a relationship. Hence, such cognitive resemblance was found to exist, supporting the claim that these bibliographically coupled documents did represent work of the same research specialty.

In Glänzel and Czerwon (1995 and 1996), it was shown that bibliographic coupling can be used to identify "hot" research topics as represented by so called "core documents", which were identified through the application of appropriate thresholds for both the number of common references as well as the strength of coupling links. Using the whole annual accumulation of the 1992 volume of SCI, about one percent of all documents was found to be core documents. [14] A detailed analysis of both key words in titles and indexing terms indicated the representation of important research front topics and through several expert questionings it was found that most core documents belonged to a few high impact documents of a specialty. Performing a cross-national citation analysis covering two years, Glänzel and Czerwon (1996) found further empirical evidence substantiating the claim of high impact of core documents: (1) only a small share (15.7%) of core documents were not

[14] The data comprised 511,899 articles, notes and reviews, and only the document type "letters to the editor" was excluded on grounds of generally not belonging to research fronts.

cited, (2) almost two thirds of core documents were cited above the average and (3) a relatively large share (10%) of core documents were highly cited.[15]

The method presented proceeds from the model suggested by Sen and Gan (1983) and uses the C.A. as a measure of the coupling strength. Glänzel and Czerwon restricted their analysis to a subset of coupled documents where each document was coupled with at least ten coupling links with a minimum C.A. of 0.25 to other documents. The choice of thresholds was based on both theoretical considerations and empirical findings. According to the researchers, a lesser number of coupling links could bring about that documents published in series might influence results, whereas a greater number of coupling links would eliminate smaller research topics. They also claimed that a certain filtering of noise is necessary in order to avoid less characteristic coupling links between documents and that a value of the C.A. considerably lower than the stipulated would not accommodate this need. Also, too high a value of the C.A. would dramatically diminish the number of coupling links, leading to a serious decrease of found documents.

The researchers concluded that documents connected by strong bibliographic coupling links can provide insights into the structure of research fronts and be applied for science mapping purposes. They also highlighted that bibliographic coupling has several advantages in comparison with cocitation clustering, the most important being the possibility to capture the early stages of a specialty's evolution.

2.2 Cocitation Analysis

The cocitation frequency, a measure related to bibliographic coupling, was independently introduced in 1973 by Small and Marshakova. This form of document coupling was defined as the frequency with which two documents are cited together (Small, 1973). The *cocitation strength* is then defined as the number of identical citing items. Small (ibid.) also gave a formal definition of cocitation coupling as follows:

> If A is the set of documents which cites document *a* and B is the set which cites *b*, then $A \cap B$ is the set which sites both *a* and *b*. The number of elements in $A \cap B$, that is $n(A \cap B)$, is the cocitation frequency. The relative cocitation frequency could be defined as $n(A \cap B) \div n(A \cup B)$.

When measuring cocitation strength, the degree of association between documents as perceived by the population of citing authors is measured. Hence, to be strongly co-cited, a large number of authors must cite two earlier works. Small argued that due to the dependence on authors, cocitation patterns can change over time, just as vocabulary co-occurrences can change over time as subject fields evolve. Furthermore, Small noted that bibliographic coupling

[15] For the assessment of the impact of core documents published during 1992, citations received during the period 1992-1993 were compared with corresponding journals' citation impacts. "Highly cited" meant that a core document had received at least five times as many citations as the average article in the journal in which it was published.

is a fixed and permanent relationship because it depends on references contained in coupled documents. That is, once two documents are published, their coupling is established through their references, whereas the cocitation strength between any two documents will vary over time. Small argued that frequently cited documents could be assumed to represent key concepts, methods or experiments in a field and that cocitation coupling could indicate relationships between such documents.

To illustrate his ideas, Small empirically tested how cocitation patterns would develop in a set of highly cited particle physics documents from the first quarter of the 1971 SCI. Setting a threshold of nine citations, all references in documents citing a highly cited fixed document were collected in one cluster and the cocitation relations in this cluster were analyzed. Ten important documents finally constituted a cocitation network which was analyzed on a detailed level both as to different citation relations as well as to content. Small found no clear relationship between bibliographic coupling strength and cocitation frequency, but direct citations and cocitation seemed to correlate. In several instances, highly co-cited documents were not bibliographically coupled at all. Small suggested that these results indicated that bibliographic coupling should be a less reliable indication of subject relatedness than cocitation coupling. Small concluded "[t]hat an interpretation of the significance of strong cocitation links must rely both on the notion of subject similarity and on the association or co-occurrence of ideas".

Small assumed the usefulness of the following two information retrieval applications based on cocitation couplings:

i. a secondary index based on highly co-cited documents which would allow sequential searches through a citation index; and

ii. the creation of a cluster or core of earlier literature for a particular specialty, serving as a basis of an SDI system[16].

Small also foresaw the use of cocitation in the study of the specialty structure of science and as a way to monitor the development of scientific fields and the assessment of interrelationships within and between specialties.

In another article in the following year, Small and Griffith (1974) reported an experiment where a computer based system was used to identify clusters of highly interactive documents in science. This experiment aimed at a technique that would make it possible to explore the entire structure of specialties and their relations. The authors argued that an overall view of discipline structure could be obtained by analyzing citation patterns between journals, but for the purpose of revealing the fine structure of science, this was too broad a unit. Partly based on evidence that the onset of rapid specialty growth is accompanied by the emergence of key documents which are quickly and

[16] SDI (Selective Dissemination of Information) is a current awareness service providing researchers with current and pertinent publications on a specified research topic.

frequently cited, the researchers' hypothesis was that the specialty structure could be revealed by clustering frequently cited documents.

In this investigation, three percent of one quarter of the annual 1972 SCI file was applied, and those specialties, which were most active during the early part of 1972, were identified. Using the single link cluster method, clusters were formed at different levels of cocitation strength and presented as graphs.[17] A central question was whether clusters correspond to identifiable subject matter specialties. Small and Griffith mentioned several possible errors, namely:

- specialties which are not socially or intellectually related to one another could be linked together;
- the clusters themselves could be fragments of single specialties; and
- clusters could bear little or no relationship at all to the specialties.

The evaluation of their method was pursued by inspection of clusters at different levels. Known specialties like "particle physics", "nuclear physics", "reverse transcription" and "Australian antigen" were immediately recognized, though at different levels of coupling strength. Connections by cocitation between documents in clusters were also inspected and recognized as valid trails of history of research in some specialties. The researchers found that cocitation links between clusters at different levels generally linked to appropriate clusters. They also tried to examine word usage of titles in documents citing documents in clusters. Moving through a particular network, discontinuation of a vocabulary would indicate failure to group together documents of a certain specialty. On the other hand, identical or very similar use of vocabulary but the absence of cocitation links would indicate failure to group documents of a certain specialty. With a point of departure in these assumptions, Small and Griffith examined the two physics' clusters by creating word profiles for each of the cited documents. Such a word profile would be constituted by the four most frequent title words from all documents citing a particular document. The researchers found that the vocabulary within the clusters was consistent with their perceived subject content. They argued that the very existence of document clusters which, by definition, have a high degree of internal linkage, should be strong evidence for the specialty hypothesis. They also stressed that specialties of science did not seem to be isolated from one another but connected by weak links as almost all documents in their sample were linked, although tenuously. Their final conclusion was that science and its literature could be conceived as a network of specialties, each specialty being the centre of an interactive and intense communication system.

[17] Sub-files of cocited pairs were created at four levels, i.e., level 1, level 3, level 6 and level 10. Level 1 comprised all generated pairs of cocited documents, level 3 comprised pairs cocited at least three times, level 6 comprised pairs cocited at least 6 times and level 10 comprised pairs cocited at least 10 times.

In a sequel to the above article (Griffith, Small, Stonehill & Dey, 1974), Griffith, Small and their fellow researchers reported their attempts to create maps on different scales of scientific literature by applying the same set of bibliographic records. The intention was to create an overview of all highly cited documents in natural science, to reflect a single specialty in detail and to present a new clustering technique, namely, cluster cocitation. Applying this technique, weaker links of co-cited document pairs connecting any two clusters on a cocitation threshold lower than stipulated for the inclusion of co-cited documents in clusters were used. Hence, cluster cocitation is a count of the number of times that documents in two different clusters have been cited together. A preliminary map of science, constituted by interconnected clusters was produced and it was found that a majority of clusters were small, containing three to four documents, while a few - biomedicine, chemistry and physics - were of considerable size, being the major poles.

Regarding the phenomenon of small clusters, the researchers suggested that these clusters might be fragments of larger clusters that would not emerge at the level of applied thresholds. At several points, the researchers encountered macro clusters which could not be analyzed as single specialties, e.g. "cancer research – reverse transcription". Taking this cluster as an example (similar results were reached from analyzing the remaining macro clusters), the researchers analyzed the distribution of strengths of links within and between sub-clusters, i.e. groupings within macro-clusters. They found that about half of all possible links[18] between sub-clusters were absent and a negligible number of links exceeded an average of two cocitations. In addition, only a small number of sub-clusters with internal linkages of an average cocitation strength less than three cocitations were found. This led the researchers to conclude that the structure of macro clusters could be reduced to the following two components:

i. the internal structure within small groupings of documents; and

ii. a structure of few linkages greater than zero which hold the smaller groupings together.

In the above-mentioned experiments, (Small & Griffith, 1974; Griffith, Small, Stonehill & Dey, 1974) integer citation thresholds were used to select highly cited documents, and cocitation was defined on an integer basis. Also, clustering thresholds were set in terms of the integer cocitation frequency. However, these early approaches presented some difficulties which include:

- very highly cited documents, such as biomedical methodology documents had to be removed in order to prevent the creation of very large macro-clusters;

[18] The number of possible links refers to the number of 2-combinations of a set of n distinct documents, i.e. $C(n, r) = C(n, 2)$ or $\binom{n}{2}$. See equation 2.1 under Section 3 in Chapter 2.

- integer cocitation counts introduced a size dependency, i.e. highly cited documents also tended to be highly co-cited, which biased analyses against smaller research areas; and

- it is well known that there are differences as to the length of reference lists between disciplines, e.g. between mathematics and biomedicine. When an annual slice of science is analyzed, a higher reference intensity per document could affect the cocitation clustering in the following two ways:

  i. it increases the number and proportion of items from the discipline that has a higher reference intensity per document; and

  ii. it increases the strength and density of cocitation links formed amongst documents from the discipline that has higher reference intensity per document.

Consequently, in 1975, cocitation normalization was introduced to overcome some of these problems (Small & Sweeney, 1985). Applying either the Jaccard coefficient or the cosine function for normalization, one could partially overcome the problems of highly cited method documents and size dependency. The "Jaccard coefficient" (commonly referred to as the Jaccard's index) is a well-known measure of the similarity *S* between two objects A and B, which counts the number of common attributes divided by the number of attributes possessed by at least one of the two objects:

$$S_{A,B} = \frac{|A \cap B|}{|A \cup B|}, \tag{2.4}$$

In the context of cocitation analysis, this function is expressed as:

$$NCS_{ij} = \frac{C_{ij}}{(C_i + C_j - C_{ij})}, \tag{2.5}$$

and the cosine function (commonly referred to as Salton's cosine formula) is expressed as:

$$NCS_{ij} = \frac{C_{ij}}{\sqrt{(C_i \cdot C_j)}}, \tag{2.6}$$

where:

NCS_{ij} = the normalized coupling strength between document *i* and *j*;

C_{ij} = the number of cocitations of document *i* and *j*;

C_i = the number of citations of document *i*; and

C_j = the number of citations of document j.

Both measures take values in the interval [0,1].

Based on experiments on the 1979 volume of the SCI, Small and Sweeney (ibid.) proposed the following improvements to the cocitation cluster technique:

i. fractional citation counting; and

ii. variable level clustering, with a maximum cluster size limit.

The first step in the cocitation cluster method is to set a threshold for the minimum number of citations a document needs to receive in order to participate in the clustering. Using fractional citation counting, each reference is assigned a weight corresponding to the length of the reference list, e.g., if a reference list has the length of ten items, then each item is assigned 1/10. This procedure generally has the effect of giving documents with short reference lists greater weight relative to documents with longer ones. Hence, some of the problems concerning the bias toward high referencing fields should be avoided using fractional counts. It was also found that fractional counting increased the range of subject matters covered by clusters.

As the optimum cocitation level from the standpoint of recall and precision varies from specialty to specialty, the difficulty lies in the selection of the best cluster version. In order to deal with this problem, a strategy of variable level clustering, where clusters could be generated at different thresholds, was used. Using maximum size as a limiting parameter, a cluster would be generated at the lowest possible cocitation strength level, provided it did not exceed the specified cluster size. If so, the program would increment the cocitation level and try to form a cluster again on a higher level. This strategy breaks large clusters into smaller fragments, but since it allows the initial cocitation threshold to be set lower, it also allows for smaller clusters to become larger. The conclusion was that both fractional citation counting and variable level clustering improved the results when cluster analysis was applied on an interdisciplinary database like the SCI.

The methodology developed by Small and allies have been seriously criticized by Leydesdorff both on methodological and on theoretical grounds (1987). Leydesdorff criticized the choice of methods preceding the model building and the exclusive focusing on the validation of the outcomes on behalf of the validation of methods. Leydesdorff meant that based on ad hoc hypotheses, which were basically wrong, Small and co-workers had assumed that "[t]he very existence of document clusters/.../is strong evidence for the specialty hypothesis" (Small and Griffith, 1974). Leydesdorff argued that this was a fallacious argument as cluster analysis always generates a cluster structure and that the real question was to determine what the structure represents. The methodological decisions were further criticized with regard to the choice of the single link cluster method, on grounds of not generating results consistent

with results obtained by other analytical techniques.[19] The single link method is also known to produce loosely bound clusters and Leydesdorff suggested that results derived from the application of this method might well be artefacts of the method and not reflect the structure of science.

From a research policy point of view, the concept of cocitation cluster analysis was also heavily criticized by Oberski (1988), where one of several points of criticism was directed to the statistical instability resulting from both the application of the single link cluster method and the arbitrary application of threshold settings. Oberski concluded that "[it] remains unclear how one could possibly distinguish between perhaps real effects from statistical effects" (ibid. p. 448).

The claim that cocitation analysis is a useful tool to map subject-matter specialties was further examined by Braam, Moed and van Raan in 1991, which have developed a method using quantitative analysis of content-words related to articles. Unlike the basic cocitation cluster method, here the authors investigated both documents grouped by the principle of solely citing documents of a particular cluster of co-cited documents, as well as the cocitation clusters themselves. The single link cluster method was applied for the grouping of cocited articles.

Based on findings, the authors concluded that the question if all topics covered by a data set can be identified by cocitation clustering, can only partially be answered by comparing results for different sets of thresholds of (normalized) cocitation strength as some research areas might lack a consensual referencing. Still, findings suggested that cocitation clustering does display research specialties, although these may be fragmented into several clusters. It was also found that cocitation clustering only partially revealed the literature relevant to identified research topics of the citing literature (a specialty's current work) and that interrelations between clusters seemed to correspond to cognitive relations on a higher level then research specialties.

The authors concluded that the method applied provides a useful instrument for the description and evaluation of cocitation analysis in terms of the cognitive content of clusters, cluster coherence and differentiation as well as the recall of specialties current citing articles.

[19] "Other analytical techniques" refers to multidimensional scaling, factor analytical approaches and Ward's cluster method.

3. SUMMARY AND FOUNDATION FOR THE RESEARCH DESIGN

3.1 Origination of Methods and Direction of Development

Some resemblances as well as differences are at hand when comparing bibliographic coupling with cocitation analysis. To begin with, both bibliographic coupling and cocitation analysis were originally assumed to be applied for information retrieval purposes, though cocitation analysis was additionally and originally suggested for science mapping. One can also see a partly parallel development of these methods. In both cases, the original elaborations were followed by experiments on large scales in order to be able to generalize findings, in particular to interdisciplinary contexts.[20] The issue of cognitive association between documents was then further elaborated and normalized measures for the association of documents suggested.

However, the relation between bibliographic coupling and science mapping is considerably weaker in comparison with cocitation analysis. The application of cocitation analysis in science mapping and the generation of science maps, follow a clearly discernable track with a series of connected articles. This development has not been paralleled by bibliographic coupling. Based on profound empirical findings and theoretical considerations, bibliographic coupling was highlighted as a science mapping tool first in the '90s (cf. Glänzel & Czerwon, 1995 and 1996).[21] By that time, the cocitation cluster technique had undergone several adjustments and refinements, new forms of cocitation relations had been explored and a corpus of research articles providing empirical experience was already at hand.

3.2 Comparison of Properties of Methods

A complementary aspect of bibliographic coupling when compared with cocitation cluster analysis is that the more current published research can be mapped. Hence, "[s]napshots of early stages of a specialty's evolution..." can be provided (Glänzel & Czerwon, 1996). This is so because there will always be a time lag between the current published research and the generation of a sufficient number of received citations that can facilitate stable sets of cocitation data for mapping. Hence, cocitation analysis has a clear shortcoming in comparison with bibliographic coupling with regard to topicality, as bibliographic coupling can capture new lines of research as soon as findings have been published.

Another marked difference between cocitation analysis and bibliographic coupling concerns the identification of research specialties. According to

---

[20] *Interdisciplinary* contexts meant in practice that the *mutltidisciplinary* database SCI were applied for the experiments refered to. More precisely, a multidisciplinary environment would allow for the mapping of cogntive relations between different disciplines, i.e. the mapping of interdisciplinary researches.

[21] Still, a few research articles have applied bibliographic coupling for science mapping purposes with results that indicate that it may be be successfully applied for science mapping purposes (Sharabchiev, 1988; Persson, 1994 and Jarneving, 2001). However, the applicability of bibliographic coupling as a science mapping tool were not exhaustively elaborated in neither of these articles.

Small's theory of cocitation analysis, research specialties are identified through the cocitation of *highly cited* papers. These papers are then seen as key-documents of a specialty (Small, 1974). These key-documents are regarded as symbols or markers of important concepts and the cocitation of such documents is then the measure of association or co-occurrence of ideas (Small, 1973; 1977). Hence, "cocitation identifies relationships between papers which are regarded as important by authors in the specialty, but which are not identified by such techniques as bibliographic coupling" (Small, 1974). The clusters generated on basis of the grouping of such cocited papers should then be representations of scientific specialties (ibid.). With regard to bibliographic coupling, claims can generally not be made that documents represent key-concepts or have a central meaning to researchers of the field under study. Hence, applying cluster analysis based on bibliographic coupling, one could not make the same claim of identifying the cognitive core of a research specialty, i.e., generally, there would be no selection criteria for the identification of significant articles.[22] Findings concerning "core documents" may point to an exception, as these have indicated that these often are high impact papers of specialties (cf. Glänzel & Czerwon, 1996). One could, however, assume that more empirical research is needed to elucidate the relation between the criteria of core documents and citation impact. Therefore, and as for now, a perhaps more justified assumption would be that current research themes, rather than the cognitive structures of specialties, may be mapped through cluster analysis based on bibliographic coupling. Weather such research themes would reflect core issues of a specialty or more peripheral aspects may perhaps be reflected by subsequent citations. As the cocitation cluster approach has rendered some severe critique (e.g. Leydesdorff, 1987; Oberski, 1988), the question whether this method in fact mirrors the specialty structure of science may not be conclusively solved. It should however be clear that both methods, cocitation analysis and bibliographic coupling, have the ability to group cognitively related documents, hence their significance for scientific information provision is obvious.

3.3 Presumed General Problems of Citation Based Document Mapping

It seems reasonable to assume that some general problems exist comprising both cocitation clustering and bibliographic coupling clustering. On the basis of previous findings from cocitation clustering and theoretical considerations, a number of problems including their inherent relation to one another seem evident. They are related to:

i. cluster size ;

ii. threshold settings;

iii. fragmentation;

[22] However, nothing hinders that only coupling relations based on the more cited references are applied for the establishment of associations between source articles, as was done in Peters, Braam & van Raan, 1995. However, the appropriateness of this approach as a standard procedure could seriously be questioned.

iv. coverage and of topics; and

v. basic assumptions.

Concerning cluster size (i), the problems encountered in cocitation analysis was for all the creation of macro-clusters, which may be considered as an effect of the applied cluster method, which consistently has been the single link method, and several adjustments of the cocitation cluster method have subsequently been considered necessary. The partition of sets of documents by cocitation clustering has commonly also led to the generation of a large number of singleton clusters and smaller sized clusters, the number of which increases by the raising of coupling threshold. As a too large number of clusters in itself would introduce noise and hinder any intelligible and comprehensive analysis of structural aspects, smaller sized clusters are usually excluded in the analysis, if at all accounted for. The issue of cluster size is related to both (ii) and (iii).

The purpose of setting thresholds (ii) in cocitation clustering has aimed at the filtering out of noise and identification of the more significant representatives (the more cited documents) of research specialties, (Small, 1973; Griffith et al, 1974). The basic difference between cocitation clustering and clustering of bibliographic coupled articles with regard to threshold settings is that in the former case, aggregations of citations guide the selection of items that should participate in the analysis. Set thresholds of cocitation strength may additionally restrict the original population of documents. In the case of bibliographic coupling, the selection of articles is based on the strength of similarity between objects, that is, articles lacking significant associations to other documents are filtered out. One may additionally apply thresholds regarding the number of links between bibliographically coupled articles at a certain minimum coupling strength (or normalized coupling strength) and in this way filter out articles that are less central in the network of interrelated articles. The setting of an "appropriate" cocitation coupling threshold for a particular specialty is, however, difficult and heuristic methods are usually applied (Small, 1977).[23] With the exception for the strict criteria of "core documents", the problem of coupling threshold addressed by Small likewise applies to bibliographic coupling.

Related to the issues of threshold setting and cluster size is the question of fragmentation (iii). Findings from research in cocitation analysis have suggested that cocitation clustering does display research specialties, although these may be fragmented into several clusters (Braam, Moed & van Raan, 1991). An assumption could be made that the cause of fragmentation sometimes could be due to a too severe threshold setting, but also, one could

[23] Applying the single link cluster method, an attempt to establish intervals of normalized coupling strength in which stable cluster sizes occur and the generation of macro clusters is avoided has been made (Braam, Moed & van Raan, 1988). However, it seems rational to separate the question of "noise and signal", from the question of method of partition.

assume, sometimes reflect actual circumstances when a research area is split up in diverging directions of research and re-modeled.

Another important issue for both cocitation analysis and bibliographic coupling should be the extent to which topics covered by a population of research articles are identified by the applied method (iv). With regard to cocitation clustering, research in this issue (Braam, Moed & van Raan, 1991) has shown that low "recall" of the "current work" of specialties is related to a lack of consensus as to the previous literature. [24] Generally, for cluster analyses based on either cocitation relations or bibliographic coupling relations, when there is a lack of consensual referencing, shares of the literature on the same topic may be lost. What actually is mapped, is a slim strip of consensus that associates a fraction of all documents of a specified field under investigation (the selected document population), which subsequently is partitioned in subsets (clusters) which are claimed to represent research specialties in the case of cocitation cluster analysis. Hence, the exhaustiveness of the mapping of a research specialty is usually unknown when applying citation based mapping methods.

Lastly, with regard to (v), the basic assumptions should be nearly the same for both cocitation cluster analysis and the proposed method based on bibliographic coupling. In short, the cognitive association between the citing document and the cited document should be based on the use of the citing document, and the selection of cited documents should not be influenced by randomness. As current mapping techniques do not usually assign weights to a cited reference,[25] the equal treatment of cited references may be considered a problem leading to a loss of precision, should it be motivated to establish the cognitive relation between a citing and a cited document. A difference between cocitation analysis and bibliographic coupling is that in bibliographic coupling analyses no obvious selection criteria of quality exist, hence, the number of basic assumptions is somewhat reduced with regard to bibliographic coupling.

With regard to the different types of couplings, cocitation coupling and bibliographic coupling, the significance of a single coupling (one cocitation or one shared reference) should be related to how well appropriate basic assumptions hold, which in practice is not feasible to establish.

3.4 Methods of Partition

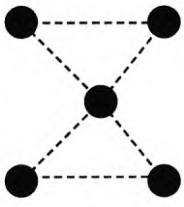
Though much of the prerequisites necessary for the development of bibliographic coupling into a practicable mapping tool are at hand through previous research (for all the establishment of cognitive relations between coupled documents), there is a lack of empirical experience concerning principles of partition of bibliographically coupled document populations. As

[24] A low recall in this meaning means that only a smaller share of source articles that are semantically similar direct references to the cocitation cluster that represent a research theme that is common for these source articles.

[25] This is to some extent a technical question and one could imagine that the frequency of references in a fulltext to a document could be considered.

previously described, the cocitation cluster analytical method has been criticized on several grounds of which an important one is the choice of cluster method. Undesirable effects like the chaining phenomenon may lead to less coherent clusters, especially when less homogenous populations are analyzed (interdisciplinary research settings) due to spurious links between documents. Therefore, as discussed in Sub-section 1.4.2 in this chapter, the idea of applying a cluster method that would not have this drawback is appealing. Certainly, other problems may be introduced, but the testing of a cluster method that could be expected to generate more coherent clusters seems motivated. The idea of such coherent groupings of documents was already put forward by Kessler in 1961 as "criterion B", where an interrelated group of documents is a group where every member has at least one coupling unit with every other member of the group. Likewise, the notion of "cliques" of bibliographically coupled documents, suggest this type of grouping of interrelated documents (Sen & Gan, 1983). The criterion stated for the generation of G_B groups (Kessler, 1962) and subsequently the notion of "bibliographic cliques" (Sen & Gan, 1983), are, however, insufficient from an algorithmic/practical point of view. Though the condition to fulfill for the generation of G_B groups applies to the notion of complete graphs or "cliques", this condition could be fulfilled in different ways, as illustrated in Figure 2-5, where two possibilities to form a complete subgraph where $n > 2$ are indicated.

Figure 2-5: Two Ways to Form a Complete Subgraph



Note: Points denote documents and lines bibliographic coupling links.

Hence, from a practical point of view, it is necessary to decide on the most appropriate algorithm.

3.5 A Foundation for the Research Design

It can be concluded on theoretical grounds that the proposed method cannot substitute the cocitation cluster analytical method. Based on the assumption that highly cited documents represent important concepts of a specialty, claims that research specialties' cognitive structures can be mapped are made by the advocates of the cocitation mapping approach. Such claims cannot be made with regard to the proposed method as no conclusive criteria exist that would identify the key-documents of a specialty on the basis of bibliographic coupling. It seems, however, reasonable to assume that specialties would be identified by the proposed method, but it would remain unclear to what extent core issues or more peripheral issues of a specialty would be mapped. Also, the question what really may be mapped applying citation based science mapping methods in general remains not fully elucidated, which, amongst several things, concerns the question of the exhaustiveness of citation based science mapping.

In addition, with regard to both methods, problems remain with the appropriate setting of thresholds, the choice of the most appropriate cluster method and also with the relation between the two. Here, one could hope, empirical experience may by time be generated that would at least shed some light on these complex issues, but for now, more heuristic methods must be applied.

In all, it could be argued that many elements of uncertainty are attached to citation based science mapping, in particular when the objective is set to map the cognitive structures of specialties. However, should the objectives be related to scientific information provision or information sharing needs, the uncertainty attached to citation based mapping in general should have lesser importance as the topicality and relevance of obtained information should be the first priority, not the exactness nor the exhaustiveness of the mirroring of specialties cognitive structures.

A rational stand-point would be that more empirical evidence is needed if rightful claims of valid depictions of specialties' structures could be made and that this may imply complex methods where several techniques are combined (cf. Braam, Moed & van Raan, 1991).

On the basis of the aforementioned reasons regarding limitations of the proposed method and citation based science mapping in general, a balanced presumption of what could be mapped applying the proposed method is needed. Given is the ability of the proposed method to generate coherent cluster with regard to the applied measure of document similarity (bibliographic coupling) and the chosen cluster method. On basis of previous findings regarding subject similarity between bibliographically coupled documents, it could also be presumed that such coherent clusters generally would be subject coherent, given appropriate thresholds of coupling strength or normalized coupling strength. Applying the proposed method, the most current and consensual research would be reflected, though central aspects and outlines of specialties' cognitive structures would generally not be expected.

It can from this reasoning be presumed that the proposed method may serve a complementary purpose in relation to the cocitation cluster method in the context of scientific information provision. Hence, the identification of current and coherent research themes (rather than specialty structures) as mirrored by reciprocally subject related documents should be the optimal outcome expected when applying the proposed method. The desirable ability of bibliographic coupling applications to map the most current published research should be contrasted with the retrograde mappings of the cocitation approach. Bibliographic coupling methods may therefore provide scientists with more current and valuable information. In comparison with the application of bibliographic coupling for searching and retrieval of related documents, the assumed capability of the proposed method to generate subject coherent groups of documents would mean a considerable progress in comparison with the retrieval of a number of rank ordered bibliographically coupled documents as more information would be obtained.[26]

[26] The ISI search and retrieval facility Web of Science is an example where bibliographic coupling is applied for retrieval purposes. The output is a ranked list of documents bibliographically coupled with a test paper (P_o), using Kessler's original expression. In comparision with a ranked list, clusters of articles would in addition contain the information inherent in the links between the articles and also provide an over view of current research themes.

CHAPTER 3: RATIONALE AND RESEARCH DESIGN

Based on the theoretical foundation presented in the previous Chapter 2, the rationale for the research design and the questions that it addresses are given in this chapter.

1. RESEARCH SETTINGS

In order to reach a comprehensive understanding of the proposed method's applicability as a science mapping tool, it was decided that it be tested under different environments. This is motivated by the fact that citation behavior, as reflected by lengths of reference lists and time-lag between publication and citation, differs between different fields (cf. Small & Sweeney, 1985), which affect the strength, number of bibliographic coupling links and the density of the citation network. Though the precise choice of fields was of no immediate importance, a variation with regard to size of fields (publication output), referencing character and subject matter was strived for. It was also deemed important to test the proposed method for core document mapping. This is motivated by findings in previous researches where the important role of core documents in the science communication system has been established. This implies a large scale multidisciplinary research setting as the incidence of core documents generally should be low when delimited to a single field. From another viewpoint, it is also of interest to evaluate the proposed method in a research setting that is not restricted in terms of discipline borders (cf. Weinberg, 1974 and Vladutz & Cook, 1984). Such a research setting would give access to an immensely larger network, covering also cognitive relations transcending discipline borders. Based on the aforesaid reasons, it was decided that empirical tests be carried out in four different research settings. Each of the first three corresponds to a certain field of research. The three fields were:

i. scientometrics;

ii. organic chemistry; and

iii. pure and applied mathematics.

The fourth setting is on a multidisciplinary basis comprising an annual volume of the SCI. For convenience, henceforth, these four research settings are referred to as cases, Case 1 to Case 4.

2. RATIONALE AND RESEARCH QUESTIONS

A basic assumption underlying the design of this study is that cluster analysis based on bibliographic coupling can not identify the cognitive cores of research specialties in the same direct fashion as cocitation cluster analysis, as there generally exists no clear indicator of document impact for currently published research articles. It is assumed, however, that clusters based on bibliographic coupling do reflect research specialties, though the extent to which the core of a specialty is identified can not be established on basis of the aforesaid reason. Therefore, the term "current research themes" is regarded a more proper expression of what clusters in fact may reflect. This also brings

about that the design of this study is delimited to establish if the proposed method is capable of identifying such current research themes, rather than capable of elaborating the cognitive structures of specific specialties. By way of introduction two main purposes for science mapping were mentioned: (1) the study of the specialty structure of science and (2) scientific information provision. Without excluding (1) as a proper area of study with regard to cluster analysis based on bibliographic coupling, the design of this study aims at the evaluation of the proposed method in the context of scientific information provision.

The evaluation of the proposed method has somewhat different aims with regard to the four cases. As for Cases 1 to 3, the objective is to test the proposed method's applicability as a tool for the mapping of a particular field of research. In Case 4, the objective is to test the proposed method in a multidisciplinary research setting with a specific focus on the mapping of core documents.

The rationales and research questions pertaining to Cases 1 to 3 and to Case 4 are presented in two separate sub-sections.

2.1 Cases 1 to 3

There are two major aspects in the evaluation of the proposed method with regard to Cases 1 to 3, namely, the relevance of cluster compositions and the agreement between intellectual-manual partitions of article populations performed by field experts and partitions generated by the complete link cluster method. The field experts' partitions will be considered as an external point of reference in this study. Should both partitions generally agree, then something similar to an expert system[27] would have been accomplished. However, little new information would be generated. On the other hand, should there be a general disagreement between two partitions, it would either indicate a failure to generate relevant clusters that are cognitively coherent, or, the adding of new information in terms of different but generally relevant clusters. The relevance of cluster compositions will be assessed by field experts where the number of documents that are not in line with the identified research theme of a cluster is established by way of inspection of articles in clusters.

The agreement between partitions generated by the complete link cluster method and the intellectual-manual method will be assessed through pair-wise comparisons of partitions. The following variables will be compared:

i. the concentration of articles to clusters;

ii. the internal coherence of clusters; and

iii. the external isolation of clusters.

[27] A computer program that performs a task that would, otherwise, be performed by a human expert.

Concerning (i), the extent to which a partition leads to a dispersion of articles to many clusters or a concentration of articles to a few clusters, may reflect fragmentation as well as the amalgamation of research themes. With regard to (ii) and (iii), the internal coherence and the external isolation of a cluster reflect the extent to which a cluster is consistent and demarcated with regard to the definition of similarity applied for the merging of articles to clusters. Deviations with regard to (ii) and (iii) between partitions generated by the complete link cluster method and field experts respectively may reflect that semantic relations between documents (as perceived by field experts) are not mirrored by consensual referencing, or, that the field experts' perceptions of cognitive resemblance between documents rely on intellectual classification schemes that supersedes both consensual referencing and semantic similarity. However, primarily the nature of the deviation between the complete link cluster method and the expert clustering is explored, not its causes.

Deviations between partitions will also be assessed with regard to the composition of clusters. Here, a qualitative approach will be applied where the dispersion of articles over clusters will be studied by visual inspection.

The research design covering Cases 1 to 3 aims to answer the following questions:

Q1. To what extent does the proposed method generate relevant clusters?

Q2. What is the nature and extent to which results generated by the proposed method deviate from results generated by intellectual-manual partitions performed by field experts?

Q3. What are the effects of the application of the proposed method with regard to applied thresholds and method of partition on document populations?

Q4. What are the implications of the results in Q1 to Q3 with regard to the application of the proposed method as a tool for the mapping of science fields?

2.2 Case 4

There are three factors which motivated the research design of this case. They are:

i. the incidence of core documents;

ii. the properties of core documents; and

iii. the properties of the complete link cluster method.

With regard to (i), as was mentioned earlier, the incidence of core documents should generally be low for a single field. Glänzel and Czerwon (1996) found that less then one percent of all items (4,534 documents) in the 1992 volume

of SCI were core documents. These were dispersed over 42 sub-fields and assigned a total of 128 journal subject categories. This dispersion of core documents over a large number of fields and specialties underlines the necessity of a multidisciplinary research setting.

With regard to points (ii) and (iii), considering the severe rule for merging of objects when the complete link cluster method is applied (see Sub-sections 1.4.1 and 1.4.2 in Chapter 2) and the role of core documents as central nodes in networks of bibliographically coupled articles with many and strong links to other articles, one could on theoretical grounds presume that core document clusters frequently would be parts of larger groups of related articles. In order to further elaborate the implications of this presumption, a strategy of mapping was outlined and will be applied.

The strategy has its point of departure in a set of clusters generated by a first partition of the population of core documents. Here, only strong links will be used for the clustering of core documents. This partition forms a base line from which two lines of mapping will be pursued:

i. In the first line of mapping, all significant (strong) links connecting core documents in clusters with any other core document will be mapped. This will result in a depiction of all significant artificially broken links between core documents in a cluster and core documents extrinsic to that cluster. The rationale for carrying out this line of mapping is that it will enable one to measure the extent of fragmentation of research themes the application of the proposed method may give rise to.

ii. The second line of mapping involves the application of links between clusters only. They will be used to successively merge clusters on two subsequent levels of fusion, where the first generation of clusters are considered objects for a second clustering, and the second generation of clusters will give rise to a final cluster fusion. The rationale for carrying out this second line of mapping is that larger specialties with complex internal structures may be mapped when the information in links between clusters is applied.

The impact of iterated clustering will be regarded with respect to the overall cluster structure, with a starting point at the base line. Changes of cluster composition on the three levels will be evaluated with regard to the following variables:

i. the internal coherence of clusters;

ii the external isolation of clusters;

iii. the reduction of the number of clusters;

iv. the increment of cluster sizes;

v. the number of isolated clusters; and

vi. the number of singleton clusters.

Concerning points (i) and (ii), the internal coherence and the external isolation of a cluster reflect the extent to which a cluster is consistent and demarcated with regard to the definition of similarity applied for the merging of articles to clusters. Points (iii) to (vi) would reflect effects that a´ priori could be expected when applying iterated clustering.

With regard to the multidisciplinary aspect of this research setting, a comprehensive expert evaluation of cluster relevance would be impracticable. Therefore, the assessment of cluster relevance (i.e. cluster subject coherence) will basically be grounded on statistical assessment of cluster properties. At the base line (first clustering), clusters are assumed to be subject consistent. This was deemed reasonable on the following grounds:

i. Previous researches (e.g. Vladutz & Cook, 1984; Peters, Braam & van Raan, 1995) have shown that strong bibliographic coupling links between research articles generally imply subject relatedness.

ii. Only strong links between core documents will be applied in the clustering.

iii. The use of the complete link cluster method will exclusively generate completely interconnected clusters.

iv. In consideration of (i) to (iii) above, subject coherent clusters should be expected.

It is therefore, presumed that changes of cluster coherence will generally mirror changes of cluster subject coherence. Likewise, changes of the external isolation are presumed to mirror the continuation or discontinuation of a specialty over levels of cluster fusion.

In order to complement findings, four cases of iterated clusterings will be presented to field experts, who will be invited to evaluate and comment on the subject coherence and separation of clusters in terms of cluster relevance on different levels of cluster fusion. The selection of these cases is aimed at finding examples from the dominant scientific fields, namely *physics*, *chemistry* and *bio-medical sciences*.

The design of Case 4 attempts to answer the following questions:

Q1. To what extent does the proposed method impose a fragmentation of specialties, when applied for core document mapping?

Q2. What is the impact of iterated clustering on the overall cluster structure?

Q3. Is there an optimal level of cluster fusion?

Q4. What are the implications of the results in Q1 to Q3 with regard to the application of the proposed method on core document data?

CHAPTER 4: METHODS AND DATA

This chapter is divided into three sections. The first section is about the basic components of the proposed method which deals with measures of document similarity, the application of the complete link cluster method, the motives for the causal use of the between groups average cluster method and a minor experimental comparison of three agglomerative hierarchical methods. The next section describes the methods used in the evaluation of the proposed mapping method. The last section presents the process of data selection, properties of the document populations in the four research settings and a discussion of threshold settings and periods of observation. The latter is supplemented with a minor experiment.

1. THE BASIC COMPONENTS OF THE PROPOSED METHOD

1.1 Measurement of Proximity

The first task when the objective is to partition sets of objects for mapping purposes is to find a method for deciding the proximity between objects. A *proximity value* is a number which indicates how similar or dissimilar two objects are. Proximity measures could be of two types as follows:

i. **Similarity measure** – a function that maps the association between two objects so that the stronger the association, the higher the number; and

ii. **Dissimilarity measure** – a function that maps the association between two objects so that the stronger the association, the lower the number.

In the context of cluster analysis, the notions of similarity and dissimilarity have a correspondence to *distances*. When applying a proximity measure that is a similarity measure, a high number corresponds to a small distance. A dissimilarity measure, on the other hand, yields the opposite result, i.e., a high number corresponds to a large distance.

Generally, the similarity or dissimilarity between two objects can be measured in two essentially different ways as follows:

i. **Local approach** – the direct similarity between two objects; or

ii. **Global approach** – the way the objects relate to other objects in the population studied (Ahlgren, Jarneving & Rousseau, 2003).

Where bibliometric studies are concerned, the local approach has frequently been applied in document cocitation cluster analysis, e.g. in the original works by ISI researchers (e.g. Small, 1973; Small & Griffith, 1974; Griffith, Small, Stonehill & Dey, 1974; Small & Griffith, 1983; Small & Sweeney, 1985) and also in research in bibliographic coupling (e.g. Sen & Gan, 1983; Vladutz & Cook, 1984). In other instances, the global approach has been the prevalent approach, as in author cocitation analysis, where the common method is to

compare an ordered pair of vectors of author cocitations and calculate Pearson's *r* between these vectors (e.g. White & Griffith, 1981; McCain, 1990; White & McCain, 1998). Also, in many applications, objects are assumed to be represented as points in Euclidean space.

The problem with the global approach is that sometimes the underlying data (e.g. values of bibliographic coupling strength or cocitation coupling strength) include associations of the type "co-absence". "Co-absence" refers to pairs of objects which might be seen as similar in the sense that both objects lack an association to other objects in the population of study. This might in some instances correctly reflect the similarity between two such objects, and in other instances, it might not. Thus, one must ask if "co-absence" contain useful information about the similarity between two objects (Everitt, Landau & Leese, 2001, p. 36). In Leydesdorff (1987), the author points out that a large amount of "missing values" (zeros in a data matrix) could be potentially problematic when applying the Euclidean metrics to citation data:

> Since missing values do not add to the Euclidean distance between two cases, those cases with large amount of missing values end up with small distances among them, and when this is the clustering criterion (as for example in the single link clustering) clustering starts at this end.

Hence, the possibility to cluster two objects on the ground of their difference from the rest of the set of objects, rather than on the ground of their similarity with each other is obvious. In this study, when performing some preliminary tests where Euclidean distances were used as input data to different cluster routines, this frequently brought about early fusions of documents on the basis of a small or zero distance in combination with a low coupling strength or the complete absence of couplings. This outcome was also found when Pearson's *r* was applied as a proximity measure. In spite of this, the global approach seems appealing from the perspective that more information is underlying the values of similarity or dissimilarity.

Due to this flaw of global measures, a local measure might be preferred. However, the original definition of bibliographic coupling strength (see Sub-section 2.1 in Chapter 2) may not represent the optimal measure of document similarity. Also, the fact that two articles have a reference in common is no guarantee that both articles are referring to the same piece of information in the cited article (cf. Martyn, 1964). In addition, the significance of a reference is not known, and references may differ in terms of their impact on the citing article (see the comment on point (v) in Sub-section 3.3 in Chapter 2). In spite of this, it still seems reasonable to assume that the probability of the cognitive relationship between two documents should increase by the number of common references. Moreover, the significance of a bibliographic coupling unit associating two articles should be inversely related to the combined lengths of the reference lists of both documents (Vladutz & Cook, 1984). Therefore, a function that normalizes for the length of reference lists is needed. This calls for the use of the C.A. presented in Sub-section 2.1 in Chapter 2 (equation 2.3). The C.A. has been applied by several other researchers in the

past (e.g. Sharada & Sharma, 1993; Mubeen, 1995; Glänzel & Czerwon, 1995 and 1966). For the sake of simplicity, this measure is defined here as:

$$NCS_{ij} = \frac{r_{ij}}{(n_i \cdot n_j)^{1/2}}, \qquad \textbf{(4.1)}$$

where

NCS_{ij} = the normalized coupling strength between article i and article j

r_{ij} = number of references common to both i and j

n_i = number of references in the reference list of article i

n_j = number of references in the reference list of article j

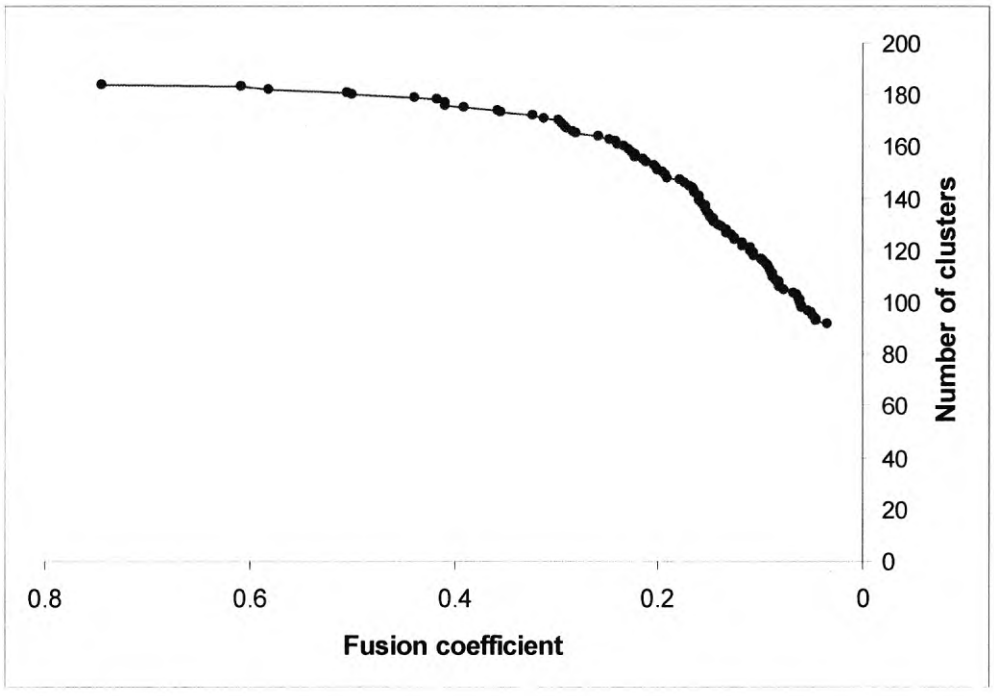
The interval is [0, 1] and $n_i = n_j = r_{ij}$ gives the maximum value.

This function will be referred to as the Normalized Coupling Strength (NCS) henceforth in this study.

1.2 Application of the Complete Link Cluster Method

Since all agglomerative hierarchical techniques reduce data to a single cluster containing all the objects, the search for an optimal number of clusters demands a decision of when to stop. Usually partitions are achieved by cutting a dendrogram at a particular height, a "best cut" (Everitt, Landau & Leese, 2001, p. 76). This requires that clear shifts of fusion levels are discernable. However, in no case did a marked hierarchical structure show up. This phenomenon is illustrated in Figure 4-1.

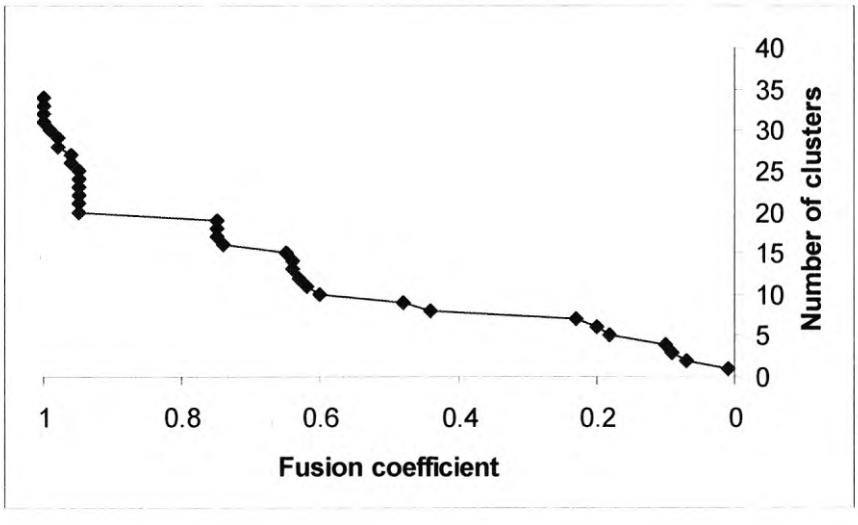
Figure 4-1: The Distribution of Clusters at Different Fusion Coefficients



Note: The graph shows the merging of 185 documents constituting the final population of Case 1. Fusions at zero level are excluded.

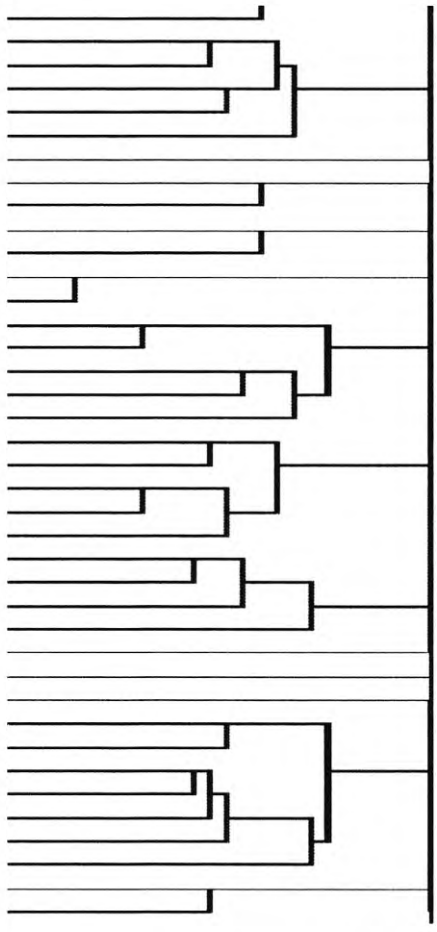
It can be seen from Figure 4.1 that the curve is essentially flat from its beginning to its midpoint, reflecting the fusion of a few clusters at higher levels of similarity. After the midpoint, the decline of the curve reflects an increasing number of mergings at considerably lower levels. This type of curve was found in each of the three cases and is more or less the opposite to a clear cluster hierarchy, where the curve initially would show the fusion of a larger share of clusters and then flatten out as the following fusions of clusters take place on lower levels, as is exemplified in Figure 4-2.

Figure 4-2: Hypothetical Curve of Cluster Fusion where there is an Inherent Hierarchical Structure in Data



Therefore, the optimal number of clusters was considered equal to the total number of clusters of maximum size clearly discernable below the fusion level of zero (below the highest rescaled distance) in the dendrogram. Generally, there was no problem to identify the optimal number of clusters in dendrograms. An example of such a "best cut" in a dendrogram is given in Figure 4-3.

Figure 4-3: Part of Dendrogram Generated by the Complete Link Cluster Method



Note: The above is a display of a part of the dendrogram from Case 2 where 268 articles were clustered. Note the typical spacing between the lower levels of fusion and the highest rescaled distance (zero level). All clusters below the highest (discernable) rescaled distance are below the "cut".

However, when the range of the scale of coefficients of similarity is wide, on the comparably more narrow "rescaled distance" in the resulting display of a dendrogram, extremely low values of similarity are sometimes not clearly separated from the zero line (zero similarity – largest distance).[28] Hence, some associations between clusters at very low values of similarity may therefore be regarded as zeros.

[28] These can be identified by the resulting agglomeration schedule in the hierarchcial clustering routines of SPSS.

When a choice of partition has been accomplished, the distribution of documents over clusters may be skewed, with a majority of clusters constituted by one or two documents. As the goal of clustering is the arrival at some kind of meaningful summation of data in a smaller number of groups of objects, a confused pattern of numerous single objects and pairs would not contribute to such a goal. Hence, in this study, clusters containing less than three documents were excluded from further analysis.

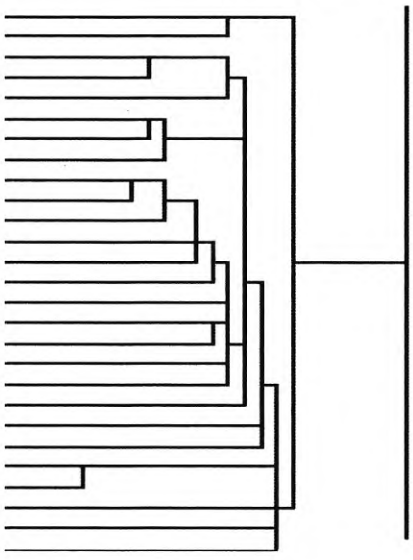
1.3 Application of the Between Groups Average Cluster Method

In order to evaluate the extent to which significant links between clusters generated by iterated clustering still remained on the last level of cluster fusion in Case 4 (see Sub-section 2.2 in Chapter 3) the between groups average cluster method was applied. The reason for this was that the complete link cluster method implied a too severe condition to be fulfilled for the fusion of clusters. With regard to the assessment of number of resulting clusters, the scale problem of dendrograms mentioned in the previous Sub-section 1.2 brought about the generation of several singleton clusters. This was, however, not considered a real problem as extremely low values of similarity do not reflect significant associations.

1.4 A Comparison of Cluster Methods

In order to substantiate theory (see Sub-section 1.4 in Chapter 2) with empirical findings, the complete- and single- link methods as well as the between groups average link method were compared over three of the research settings (Cases 1 to 3). Not unexpectedly, the single link method generated an unclear cluster structure with few large and loosely bound clusters. An example of this chaining phenomenon is given in Figure 4-4.

Figure 4-4: Example of Chaining Generated by the Single Link Method



Note: i. Connected horizontal lines designate joined documents.
ii. The part of the dendrogram shown in this figure is from the clustering of documents selected for research setting 2 – "organic chemistry" – where 268 articles made up the final population.

The application of the complete link cluster method and the between groups average link method both resulted in clearer cluster structures, though the between groups average method generally generated larger clusters and lesser singleton clusters. In a preliminary test, the complete link cluster method was further compared with the between groups average link method. Though the complete link cluster method generated clusters with a maximal density, the between groups average method sometimes generated clusters with a near maximal density (cf. Sub-section 1.4.2 in Chapter 2). Taken as an example, in the first research setting (Case 1), the average value of D for clusters was as high as 0.73, though the average coupling strength in clusters was considerably lower.[29] The extent to which the more strict rules of cluster fusion of the complete link cluster method will result in more relevant clusters in comparison with the between groups average method was assessed in a qualitative test where a field expert was invited to compare two partitions of the population of documents used in Case 1; the first being generated by the complete link cluster method and the second by the between groups average link method. A minimum cluster size of three documents was applied in both cases.

In the first partition generated by the complete link cluster method, 63 documents were distributed over 17 clusters. In the second partition generated by the between groups average link, 134 documents were distributed over 27 clusters. The field expert was asked to note any misplaced documents according to a set of rules (see Sub-section 2.1 in this chapter). A total of 39

[29] For the complete link method, the average coupling strength in clusters was 5.09 and the corresponding value for the between groups average linkage was 2.58. For a definition of average coupling strength, see Sub-section 2.2 in this chapter.

documents or 29 percent of all documents in the clusters generated by the between groups average method were regarded as misplaced whilst the corresponding figure for the complete link cluster method was six documents or ten percent. Important was the general effect on clusters' relevance this gave rise to. In the case of the complete link cluster method, two clusters, each containing three documents were regarded as irrelevant, and in the case of the between groups average method, 17 clusters were less than 100 percent accurate and six clusters were totally irrelevant. The result of the evaluation of the application of the between groups average cluster method on data from Case 1 is shown in Table 4-1.

Table 4-1: Field Expert's Evaluation of the Between Groups Average Method

Cluster	Percentage of Misplaced Documents in Clusters
1	17
3	11
4	33
6	100
10	14
13	33
20	11
22	100
23	100
28	38
29	100
30	25
31	100
33	25
34	40
35	100
36	33

Note: The table shows the results from the evaluation of those 17 clusters generated by the between groups average cluster method that contained misplaced documents.

Though more documents were grouped by the between groups average link method, the amount of low quality information was large. The field expert commented, that in his view, the large share of noise in the information made the resulting classification of documents more or less useless.

2 METHODS OF EVALUATION

2.1 The Qualitative Assessment of Cluster Compositions

The issue of cluster relevance was operationalized as the identification of common research foci of constituent articles in clusters, assessed by field experts' examination of the subject matter in articles. [30] Information concerning the subject matter of articles is contained in *Content Describing Elements* (CDE), where a CDE denotes an element in a bibliographic record which describes the content of an article in such a way that it is not easily mixed up with another (Noyons, 1999, p. 18). Such elements are document specific to a large extent, but some may be less specific for a particular article, such as author-names, journal-titles and cited references (ibid.). Titles, abstracts or publication specific key-words, supplied by the authors themselves, describe the subject content of the article and could be categorized as uncontrolled terms, whereas indexing terms externally supplied by professional indexers are controlled ones. Both can bee seen as representing problem domains or research themes (Tijssen, 1992, p.73), the main difference being that controlled terms may lack topicality, though they might sometimes be more adequate when titles are of a metaphorical type.

With regard to Cases 1 to 3, for each article in a cluster, first author names, publication years, journal titles, article titles, key-words[31] and abstracts were compiled and made available to the field expert. Field experts' examination of clusters' relevance was pursued by examination of each cluster, article by article, in order to detect inconsistencies as to subject content in clusters. Any article deviating from a common research theme of a cluster was regarded as "misplaced" and marked. A research theme of a cluster was identified when more than 50 percent of articles share a research focus. When this condition was not fulfilled, all constituent articles were counted as misplaced and the cluster regarded as irrelevant. Hence, should a tie occur, all constituent articles would be regarded as misplaced. Field experts were also interviewed and asked to comment on the extent to which any research specialties were missing.

In Case 4, field experts were presented with data on Excel spreadsheets. On these spreadsheets, titles of articles in clusters were given as well as the hierarchical structure of clusters on three levels of cluster fusion. For each cluster on the first level of clustering, any article not in line with the identified research focus of the cluster was marked. Next, the relevance of the compound cluster on the next level of fusion was assessed. When all constituent clusters on the second level of cluster fusion had been evaluated, the relevance of the merging of these clusters to the last level of cluster fusion was assessed. In this

[30] Field experts were selected on grounds of being active as researchers within fields corresponding to the assigned areas of evaluation and on holding an academic position within corresponding discipline. Personal knowledge in combination with external information obtained through University web sites guided the assemblance of a list of candidates.

[31] Here, key-words were of two types: a*uthor keywords* assigned by the author and *Keywords Plus* which are words or phrases that frequently appear in the titles of an article's references, but do not necessarily appear in the title of the article or in a list of author keywords.

way, not only the relevance in terms of misplaced articles was assessed, but also the relevance of the fusion of sub-clusters. This was deemed important as a common research theme may emerge by the fusion clusters, otherwise not clearly discernable on sub-cluster level. The field experts were asked to provide appropriate comments on these issues.

2.2 The Quantitative Assessment of Cluster Compositions

Many authors have attempted to define a cluster in terms of its internal cohesion and external isolation (Everitt, Landau & Leese, 2001, p. 6). Ideally, a cluster should therefore be internally coherent and externally well separated, meaning that it should contain articles that are reciprocally and strongly bibliographically coupled, lacking (strong) bibliographic couplings with articles in other clusters.

A measure of the internal coherence is the *Average Coupling Strength, AvgCS(C)*, for a cluster *C*. It is defined as:

$$AvgCS(C) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} CS(d_i d_j)}{\binom{n}{2}}, \qquad \textbf{(4.2)}$$

where

n = number of articles in a cluster c,

CS = number of bibliographic coupling units between two articles, d_i, d_j

and

$d_i d_j (\in C)$

Complementary to 4.2 is the density D (see equation 2.2). These two measures of cluster coherence reflect different aspects of internal cohesion and it is possible that a cluster could have an average coupling strength that is relatively high and a relatively low score of cluster density, and vice versa. The first case would occur if a cluster contained a number of articles coupled with strong links but a large share of all possible pairs of articles were not coupled. The second case, i.e., a high-density cluster with a relatively low average coupling strength would occur if all, or most articles, were coupled but with a weak coupling strength. Hence, the need for both measures was felt, as they do not substitute for one another.

In order to measure the isolation of clusters, a third measure is needed. Let C and C' be clusters of sizes k and m, respectively. *The average coupling strength between two clusters, C and C'*, *AvgCS(C, C')*, is defined as:

$$AvgCS(C,C') = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} CS(d_i,d_j)}{k \times m}, \qquad \textbf{(4.3)}$$

where

CS = number of bibliographic coupling units between two articles, d_i, d_j and $d_i \in C, d_j \in C'$

All three measures, 2.2, 4.2 and 4.3, are needed for the establishment of cluster relevance from a quantitative viewpoint. The cluster coherence provided by 2.2 and 4.2 is needed for the identification of coherent research themes, whereas the separation between clusters provided by 4.3, is needed for the identification of the discontinuation or continuation of a research theme. As there exists no external point of reference guiding claims of cluster relevance from the aspects of coherence and separation, the AvgCS(C) and the AvgCS(C, C') are better applied for the monitoring of changes and the assessment of differences of cluster coherence and isolation.[32]

2.3 Comparisons of Partitions with regard to Cases 1 to 3

The comparison of partitions has its point of departure in a set 'A' of clusters generated by the complete link cluster method. As discussed under Sub-section 1.2 in this chapter, clusters with a size less than 3 were considered as noise in the primary partition where the complete link cluster method was applied and therefore excluded from subsequent analyses. In this way, a subset 'B' is generated where $B \subseteq A$. The articles contained in subset B are then re-classified by an intellectual-manual clustering by a field expert in order to arrive at an external point of reference.

The extent and nature of deviation between the partitions generated by the complete link cluster method and the intellectual-manual partitions were assessed by a number of variables (cf. Sub-section 2.1 in Chapter 3) as follows:

i. the internal coherence of clusters;

ii. the isolation of clusters; and

iii. the concentration of articles to clusters.

With regard to (i), clusters generated by the complete link cluster method will have the default value 1.0 of D (equation 2.2), while clusters generated by field experts would have varying values. The mean or the median (depending on the shape of the distribution) of D and AvgCS(C) (equation 4.1) gives the balance point or the middle in a distribution.

[32] The density D as another measure of cluster coherence is less attached with these delimitations as it has a fixed range of 0 to 1.

With regard to (ii), the AvgCS(C, C') (equation 4.2) was applied for the measuring of distances between clusters. The mean or median AvgCS(C, C') gives the balance point or the middle in a distribution.

In order to assess the general level of interconnectedness within a set of clusters resulting from a partition, the share of the number of 2-combinations (see equation 2.1) of clusters that were coupled was calculated. As a partition may result in a number of isolated clusters, the frequency of clusters lacking any common references with every other cluster in the set of clusters from a partition, is a complementary measure.

The concentration of articles to clusters (iii) was assessed applying Pratt's measure of concentration. This measure is of general use when one wants to see how concentrated or spread out items (here articles) are when partitioned into categories (here clusters). This measure was originally suggested with the purpose of providing an index of concentration for rank-frequency distributions which permits comparisons of subject and journal concentration in various fields (Pratt, 1977). The starting point is the theoretical assumption that all articles are evenly distributed over *n* categories and the deviation from this norm is then measured. If one assumes that there are n categories and a total of t articles, in the even distribution there would be t/n articles in each category (ibid.).

Pratt's measure is given as:

$$C = \frac{2[((n+1)/2) - q]}{n-1}, \qquad \textbf{(4.4)}$$

where

C = Pratt's measure of concentration

n = number of categories

q = is the sum of rank times frequency for each category, divided by the total number of articles.

This measure will range between 0 and 1, where the most concentrated case (only one category) takes on the value of 1 and the "even" distribution the value of 0.

The application of this measure is motivated by the fact that the numbers and sizes of clusters may show a considerable variation between partitions. Though distributions of articles over clusters can be displayed on a detailed level for comparative purposes, a quantitative measure of concentration enhances the comprehension of such differences between partitions.

For the assessment of the deviation between partitions with regard to clusters' compositions of articles, the Rand index was initially used. [33] It was, however, found that this measure was less applicable as resulting values did not reflect differences between partitions with any precision. Also, contra intuitive results were arrived at when applying this method. Instead of applying a single measure for a general assessment of deviation of cluster compositions, tables showing the overlap of articles between clusters generated by the complete link cluster method and clusters generated by field experts were compiled for a qualitative assessment of deviations.

2.4 The Intellectual-Manual Partitions Generated by the Field Experts

The intellectual-manual clustering was performed by applying a card sorting method (cf. Miller, 1969; Biglan, 1973; McCain 1986), where bibliographic data from articles printed on cards were used. [34] The field experts were instructed to sort these cards into categories (piles) on the basis of similarity of the subject matter between articles and to assign a proper label to each pile. Cards were presented to the field experts without any order, and any number of piles, 1...n, where *n* is equal to the number of articles, was allowed and tentative clusters could be broken up and revised at any time.[35] Field experts were also asked to comment on the their method of partition and perception of the analyzed field's cognitive structure. It should be noted that the field experts' partitions were performed before their evaluations of clusters generated by the complete link cluster method took place, hence, their intellectual clusterings were not affected by impressions of the cluster structures generated by the complete link cluster method.

2.5 Visualization of Partitions

The distances between clusters or between articles in clusters are not comprehensible when presented as mere figures in a table. A better understanding of the pattern of distances between objects is arrived at when all distances are configured in a more than one-dimensional space. A method that is able to generate such displays of distances is MDS.

MDS could be summarized as a method for solving the problem of how to represent *n* objects geometrically by *n* points, so that the distances between points correspond to experimental dissimilarities or similarities between objects (Kruskal, 1964). By locating objects as points in a spatial configuration, one seeks to determine the theoretical meaning of this representation.

---

[33] A common method for comparing two partitions with different number of clusters is to apply the Rand Index. This index is based on the agreement between every pair of n objects, grouped by the two methods of partition being compared.

[34] For each card representing a specific article, title, abstract, key-words, author name(s), journal title and publication date were printed and formatted to optimal convenience. In addition, complete bibliographic descriptions for all articles were available when needed.

[35] In order to avoid any systematics as to the order of presentation of cards to the field expert, a third party sorted and re-sorted the initial pile of cards for every case.

Briefly, this is how MDS works. The input for the analysis should be a N·N symmetrical matrix containing proximity data. First, an object is indexed primarily by the letter i and secondarily by the letter j, and one assumes objects to run from 1 to N if there are N objects. Proximity data values connecting object i with object j are represented by δ_{ij} and distance data values between objects will be noted as d_{ij}. The central motivating concept, then, is that the distances d_{ij} should correspond to the proximities δ_{ij}, for example, by a linear function f where $f(\delta_{ij}) = d_{ij}$. As this correspondence may not be perfect, meaning a perfect monotone relationship between proximities and distances, such discrepancies, $f(\delta_{ij}) - d_{ij}$, are measured by a goodness of fit function (Kruskal & Wish, 1978, p. 24). The scaling starts with a random configuration and through a number of iterations the configuration is changed in order to find the optimal fit with the experimental proximities. Measuring how well the fitted distances match the experimental proximities, a so called "stress value" is arrived at. The stress ranges from 0-100% and a stress value of zero consequently mean that for every δ_{ij}, $f(\delta_{ij}) = d_{ij}$. A stress ranging from "excellent" to "good" is then expressed as a value from 0.025 to 0.05 inclusive, according to Kruskal (1964). Different *n* dimensional solutions are possible to choose from, but usually a two dimensional configuration is selected, given reasonable stress values.

Conclusively, MDS is a systematic procedure for obtaining a geometric configuration, or a "map", which has a certain relationship to the proximity data (Kruskal & Wish, 1978, p. 12). It is applied in this study with the intention of visualizing clusters' internal structures in two dimensional spaces. Also, by superimposing information obtained by cluster analysis on an MDS map, increased insights into the data structure can be obtained (Everitt, Landau & Leese, 2001, p.33).[36]

3. DATA SELECTION, THRESHOLD SETTING AND FEATURES OF FINAL POPULATIONS

3.1 Thresholds and Observation Period

Only significant associations between articles are sought and random occurrences, or differently put, noise disturbing the signal should be filtered away. Two variables seem to be of immediate importance. The first and most important is the coupling strength (or NCS) and the second is the number of links associating an article with other articles. The latter is foremost of importance when there is a point in selecting the more central documents in terms of positions in networks of coupled articles.

Generally, a quite severe threshold of NCS has been presumed by researchers (Sen and Gan, 1983; Glänzel & Czerwon, 1995 and 1996). This, however, implies the analysis of larger fields of science as smaller (and hence

[36] This method was found to be useful in Case 2 (See Section 2 in Chapter 6).

sometimes younger) fields may not have enough published articles to generate significant and enough many bibliographic coupling links. In Cases 1 and 3, the associations between documents are brittle, and there is no way of applying severe thresholds if a reasonable amount of articles should remain for the analysis. Here, the simplistic approach of considering a single common reference as a random occurrence was applied. In Case 2, a considerably more productive field is investigated and more strategies of threshold settings may be appropriate, depending on objectives. Here, an approach was taken which aimed at the mapping of the more central structures of consensus.

As the method of setting thresholds of coupling strength or NCS lacks a distinct method, the effect of varying thresholds could be of interest. Should more severe threshold settings be applied for Case 1 and Case 3, the effect would be a drastic diminishing of the number of articles, number of links and a moderate raising of the NCS in the original populations as can be seen in Table 4-2.[37]

Table 4-2: The Effects of Raised Thresholds of Coupling Strength in Case 1 and 3

A	Case 1			Case 3		
	No. Links	No. Articles	Md. NCS	No. Links	No. Articles	Md.NCS
0	1654	222	0.050	3476	826	0.042
1	477	185	0.148	744	579	0.089
2	207	119	0.148	323	362	0.137
3	113	86	0.190	189	233	0.178
4	65	53	0.225	115	162	0.240

Note: Column A shows the different levels of thresholds where a threshold of 1 implies a minimal bibliographic coupling strength of 2 etc.

With regard to the population of articles in Case 2, similar consequences are seen when raising the threshold of NCS (Table 4- 3).

[37] The correlation *r* between bibliographic coupling strength and *NCS* was 0.77 in Case 1 and 0.79 in Case 3.

Table 4-3. The Effects of Raised Thresholds of NCS in Case 2

Interval of NCS	No. Links	No. Articles
0.00-1.00	827544	14389
0.10-1.00	41742	11385
0.20-1.00	7777	6537
0.30-1.00	2614	3177
0.40-1.00	1057	1575
0.50-1.00	466	779
0.60-1.00	206	367
0.70-1.00	90	161
0.80-1.00	23	45
0.90-1.00	4	8

The larger population of strongly connected research articles from field of organic chemistry in Case 2 provided possibilities of setting an additional threshold of number of links in order to find articles that are more connected. The drastic effect on the number of articles on different thresholds of number of links within the interval of 0.2-1.0 of NCS is illustrated in Table 4-4.

Table 4-4. The Distribution of Articles at Different Thresholds of Number of Links within the Interval of 0.2-1.0 of NCS

A	No. Articles	Percentage	Cumulative Percentage
1	3081	48.5	48.5
2	1349	21.2	69.7
3	768	12.1	81.8
4	413	6.5	88.3
5	235	3.7	92.0
6	149	2.3	94.3
7	89	1.4	95.7
8	65	1.0	96.7
9	50	0.8	97.5
10	42	0.7	98.2

Note: Column A shows the thresholds of number of links for an article within the interval of 0.2-1.0 of NCS.

Bibliographic coupling techniques are somewhat sensitive to the length of the publication period and the relationships between articles in terms of common references should weaken as the observation period is augmented (Glänzel & Czerwon, 1996). This is so because as the distance in time between bibliographically coupled articles increases, the intersection of common references decreases due to a tendency of citing the more current documents.

Therefore, an observation period of ½ – 2 years is assumed to be appropriate (ibid).

In order to substantiate this assumption, a set of articles distributed over several years was examined. A total of 20,616 papers from the *Journal of the American Chemical Society*, distributed over a decade (1994-2003), containing a total of 815,595 references were downloaded from the SCI CDROM for further analysis. This small-scale experiment aimed at testing if there is a detectable tendency of links of bibliographic coupling to increase in strength when distances in time between articles is decreasing. In a first step, the bibliographic couplings strengths between articles were computed and then converted to coefficients of NCS. The resulting file of bibliographic coupling links was then added with the distances (publication years) between coupled articles. This file was then sorted by NCS and the distribution cut at its midpoint (the median). For the upper half of this distribution, the mean-distance in publication years[38] between coupled articles at intervals of NCS was then counted (see Table 4-5).

Table 4-5: The Distribution of Coefficients of NCS over Mean-Distances

NCS	Mean-Distance
0.30-0.40	0.50
0.25-0.29	1.00
0.20-0.24	1.29
0.15-0.19	1.34
0.10-0.14	1.67
0.05-0.09	2.13

Note: i. The distances are counted on publication years.
ii. The interval of NCS is 0.05.
iii. Md = 0.05

As can be seen from Table 4-4 above, there is a clear relation between the strength of links and the distance. Conclusively, the observation period suggested seems appropriate. For all four cases the observation periods comprised 1–2 publication years

[38] In the course of a publication (calendar) year, articles are published on different dates, hence, the maximal distance in time between two papers published during the same publication year is less but approximately a year, hence, the maximal margin of error of less but approximately a year.

3.2 Research Settings

3.2.1 Case 1

In this first research setting, the field of scientometrics is mapped. Scientometrics may be defined as the study of the measurement of scientific and technological development. Examples of areas of investigation are research evaluation, sociological phenomena associated with scientific communities (e.g. research collaboration) and comparative studies of research output on different levels of aggregation, like institutions, sectors, provinces and countries. Scientometrics overlap to a considerable extent with bibliometrics as it applies the same methods to a large extent. Scientometrics, however, may go beyond usual bibliometric techniques and other quantitative measures may be applied. Its main channel of formal communication is the journal, *Scientometrics*, which was launched in 1978, but several other journals from neighboring fields also publish scientometric articles.

The selection of data pertaining to this field was initiated by downloading the 2001 and 2002 volumes of the journal *Scientometrics* from the SSCI on CD-ROM. From this set of articles, a list of the most cited journals, excluding *Nature* and *Science*, was derived, and the four most cited journals were identified. These journals were:

i. Journal of the American Society for Information Science and Technology;

ii. Research Policy;

iii. Journal of Documentation; and

iv. Journal of Information Science.

All articles in these journals that were indexed in the 2001 and 2002 volumes of the SSCI were downloaded. As other research themes besides the scientometric ones were contained in this journal set, those articles not citing any article from the journal *Scientometrics* were filtered out, and the resulting subset was added to the set of Scientometrics articles. This was considered an acceptable strategy, considering the central position of Scientometrics. This rendered a small, but from a subject point of view, coherent set of articles.

The final set comprised 232 articles and a total of 5,548 references of which 4,272 were unique and the average number of references in articles was 24. The total number of bibliographic couplings was 3,071 connecting 222 source articles in 1,883 links.

One bibliographic coupling unit was applied as threshold of coupling strength. This approach aimed at the avoidance of random referencing, and was deemed appropriate for a noise reduction purpose.[39] After threshold settings, the

[39] In general, links with a coupling strength of one corresponded to a low *NCS*.

number of bibliographic coupling links between articles was reduced to a total of 477 containing 1,557 bibliographic coupling units. A matrix of 17,020 elements was then created on the basis of the links between a final set of 185 articles. The share of 2-combinations of articles in the matrix that were bibliographically coupled was 3 percent. After threshold setting, the median NCS was 0.15.

3.2.2 Case 2

In this second research setting, the field of organic chemistry was applied as the test arena. Organic chemistry is a sub-discipline to chemistry and concerns the study of the structure, properties, composition, reactions and synthesis of compounds that contain carbon. The structure of the carbon atom is unique among atoms and allows for a great array of compounds of importance. This gives rise to a large number of research foci and a high publication output.

The selection of data was based on a number of central journals. The identification of these journals was accomplished with the assistance of a subject specialist. The final journal set comprised the following journals:

i. Journal of the American Chemical Society;

ii. Tetrahedron Letters;

iii. Journal of Organic Chemistry; and

iv. Angewandte Chemie-International Edition.

A final journal set was compiled by downloading all articles in these journals indexed in the 2002 and 2003 volumes of the SCI CD-ROM. This resulted in an original set of 14,389 articles containing 464,106 references of which 273,513 were unique. The average number of references in the articles was 32. In this case, the considerably large number of articles and references made it feasible to identify the more central documents in the network of bibliographically coupled articles. The setting of thresholds and noise reduction was accomplished by filtering out bibliographically coupled pairs with a NCS below 0.25 from the remaining articles. This operation rendered a total of 4,368 pairs, containing 4,496 articles and 51,689 bibliographic couplings. Next, from the file containing this reduced pair list of coupled articles, only articles with at least five links to other articles were selected, which rendered a total of 294 articles. This brought about a further reduction of coupled pairs to 722 pairs. The number of articles was also reduced somewhat as 26 articles were exclusively coupled to other articles than the selected 294. In total, 268 articles containing 9,734 references made up the final set for analysis, which rendered a matrix of 35,778 elements. The share of 2-combinations of articles in the matrix that were coupled was 2 percent and the median NCS was 0.31.

A total of 49 articles fulfilled the requirements for "core documents" (cf. Sub-section 2.1 in Chapter 2), thus forming a subset possible to analyze separately.

3.2.3 Case 3

The area of investigation in Case 3 is labeled "pure and applied mathematics". The difference between pure and applied mathematics is that pure mathematics is motivated for other reasons than application, and applied mathematics concerns itself with the application of mathematical knowledge to other knowledge domains, (e.g. network analysis) and is closely related to other disciplines (e.g. computer science).

The selection of data aimed at the selection of titles of general journals on pure and applied mathematics, not specialized on any particular sub-field. This selection was accomplished with the help of a subject specialist who compiled a listing of the following seven journals:

i. Annals of Mathematics;

ii. Communications on Pure and Applied Mathematics;

iii. Inventiones Mathematicae;

iv. Journal de Mathematiques Pures et Appliquées;

v. Journal fur Die Reine und Angewandte Mathematik;

vi. Journal of the American Mathematical Society; and

vii. Mathematische Annalen.

A final journal set was compiled by downloading all articles in these journals indexed in the 2002 and 2003 volumes of the SCI on CD-ROM. This resulted in an original set of 879 articles containing 22,188 references of which 18,831 were unique and the average number of references in articles was 25. A total of 3,476 links constituted by 5,232 coupling units connected a total of 826 articles. It was decided that a threshold of one bibliographic coupling unit be set for the purpose of noise reduction.[40] This rendered a total of 744 coupled pairs giving rise to a matrix with 167,331 elements and a final set of 579 articles. The share of 2-combinations of articles in the matrix that were coupled was 2 percent and the median NCS was 0.09.

A summary of some parameters of the different populations pertaining to Cases 1 to 3 is given in Table 4-6.

[40] As in Case 1, only a few links with a coupling strength of one were assigned a high *NCS*.

Table 4-6: Parameters of the Four Populations Pertaining to Cases 1-4

Case	A	B	C	D	E
1	232	185	3	17020	0.15
2	14389	294	2	35778	0.31
3	879	579	0.4	167331	0.09

Note: Column A contains the number of documents in the original set of downloaded articles; column B the number of documents after threshold settings; column C the density of matrixes calculated as the percentage of 2-combinations of articles that were coupled; column D the size of matrixes (N elements) and column E the median NCS of matrixes. Note that all matrixes represent the final set of articles after thresholds settings.

As can be seen from Table 4-6, the most drastic reduction of articles takes place in Case 2, due to the objective of finding the more consensual structures, which is also reflected by the highest median NCS. The density of the matrix in Case 3 reflects a loose network of documents, but with a similar strength of association between articles as in Case 1. Conclusively, different research settings are presented with variations of population size, density of networks and NCS.

3.2.4 Case 4

In this case, a multidisciplinary research setting was constructed in order to identify the crop of core documents in a year's accumulation of research articles. From the SCI volume 2003 on CDROM, 619,570 items of the document type "articles" were downloaded. [41] The average number of references in a core document was 28. A total of 17,674,944 references were processed, resulting in 149,198,407 bibliographic coupling units distributed over 121,968,904 links. The number of links was next delimited to only comprise links with a NCS of ≥ 0.25, which resulted in a reduction to 267,034 links. In these links, 6,060 unique core documents were identified and constituted a final set for the analysis.

The dispersion of articles over disciplines was assessed by computing the distribution of core documents over journal subject categories assigned by the ISI. It was found that most articles were published in journals assigned more than one subject category. Hence, the topic of each journal (or article) only approximates the combined subject categories assigned to it. These compound classification codes or strings were counted and a total of 379 unique strings were found. In this sense, a total of 379 unique classification codes were assigned to the set of core documents. When counting each unique subject category, a total of 129 were found, which approximates 76 percent of all subject categories in the Journal Citation Report, Science Edition.[42]

[41] Hence, it was considered sufficient for the purpose of the study to only include items that could be categorized as genuine research articles.

[42] The JCR is a multidisciplinary journal citation database launched by the ISI, providing means for the evaluation the impact of scholarly journals on research. It covers more than 7,500 of the world's most highly cited, peer-reviewed journals in approximately 200 disciplines. It provides two editions: the Science Edition and the Social Science Edition.

The more frequent subject categories are presented with the share of core documents in which they appear in Table 4-7. As can be seen, physics dominates followed by bio-sciences.

Table 4-7: The Distribution of Core Documents over Journal Subject Categories in Case 4

Share of Core Documents	Subject Categories
11%	Physics, Applied
10%	Physics, Multidisciplinary
10%	Physics, Condensed Matter
9%	Biochemistry & Molecular Biology
7%	Physics, Particles & Fields
5%	Materials Science, Multidisciplinary
5%	Crystallography
4%	Optics
3%	Physics, Atomic, Molecular & Chemical
3%	Cell Biology
3%	Chemistry, Physical
3%	Immunology
3%	Engineering, Electrical & Electronic
3%	Cardiac & Cardiovascular Systems
3%	Oncology
3%	Endocrinology & Metabolism
2%	Surgery
2%	Hematology
2%	Multidisciplinary Sciences
2%	Physics, Mathematical
2%	Genetics & Heredity
2%	Chemistry, Multidisciplinary
2%	Neurosciences
2%	Physics, Nuclear
2%	Polymer Science

Note: i. Subject categories assigned to less than two percent of the set of core documents are not shown.
ii. In total, 129 subject categories were found.
iii. Subject categories are in most cases overlapping.

CHAPTER 5: FINDINGS

In this chapter, results from the empirical experiments are presented in accordance to case order, Case 1 to Case 4. For all cases, effects of the applied thresholds on the original populations as well as features of the same are presented in the previous Chapter 4 Section 3 in order to keep the issues of data selection, features of data, thresholds settings and features of the final document populations assembled. Hence, for each case findings are reported with a starting point in the final populations ready for clustering.

1. CASE 1: SCIENTOMETRICS

1.1 Clusters Generated by the Complete Link Cluster Method

A total of 185 articles from the original population of 232 articles were merged, resulting in 92 clusters of which 28 were singleton clusters. Approximately 34 percent of all articles were grouped in clusters containing at least three articles (see Table 5-1). These were selected for further analysis. The bibliographic descriptions of articles in these clusters are presented in Appendix 2.

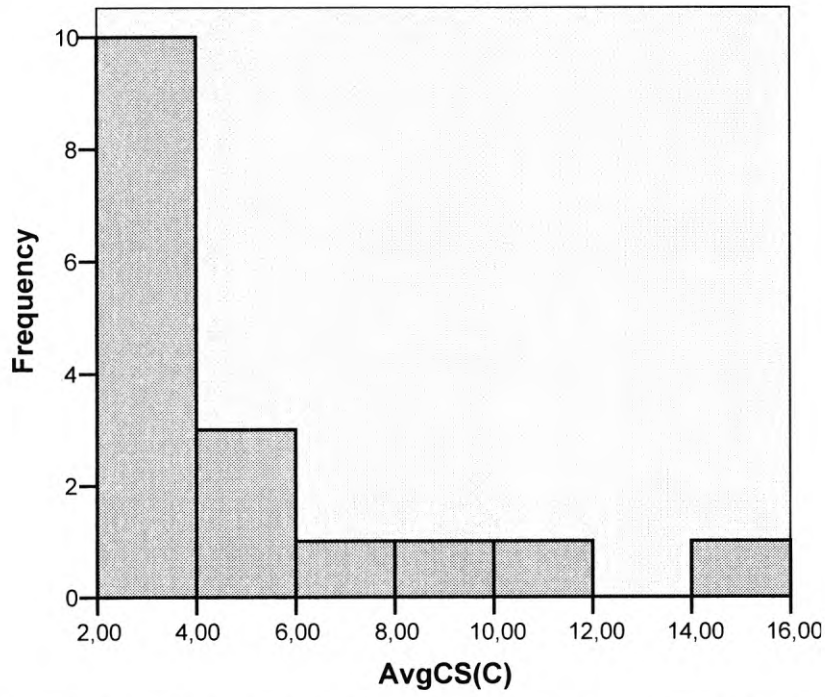
Table 5-1: The Size-Frequency Distribution of Clusters Generated by the Complete Link Cluster Method

Size	Frequency	Size · Frequency	Percentage
8	1	8	4
4	7	28	15
3	9	27	15
2	47	94	51
1	28	28	15
Sum	92	185	100

1.1.1 Coherence and Separation

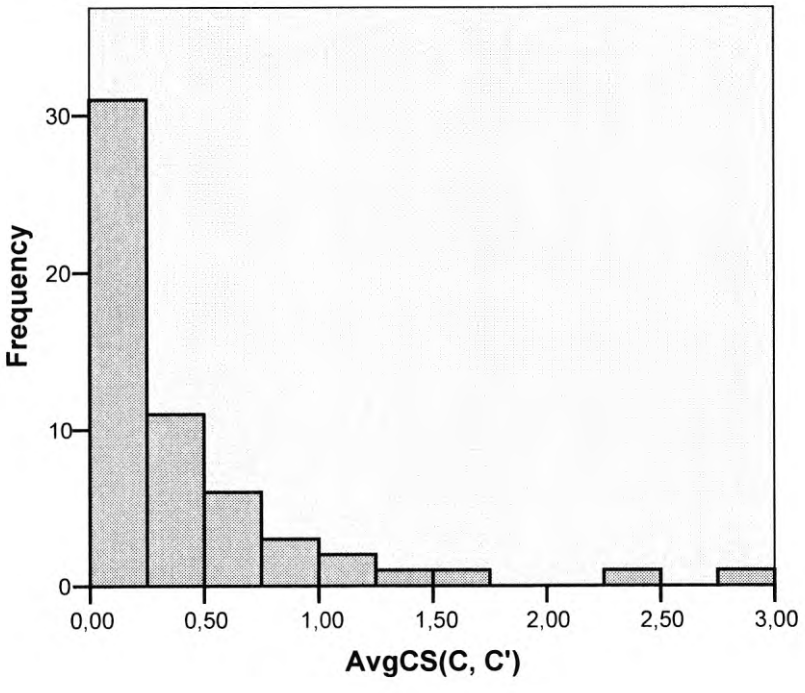
For the 17 selected clusters containing 63 articles, the AvgCS(C) (see equation 4.2) was measured. The median AvgCS(C) was 3.67 and the shape of the resulting distribution is shown in Figure 5-1.

Figure 5-1: The Distribution of Coefficients of AvgCS(C)



Next, the separation between clusters was measured as the AvgCS(C, C') (see equation 4.3). Of all 2-combinations of clusters, 42 percent were coupled and no cluster was completely isolated. The median AvgCS(C, C') was 0.19 and the shape of the resulting distribution is shown in Figure 5-2.

Figure 5-2: The Distribution of Coefficients of AvgCS(C, C')



1.2 Clusters Generated by the Field Expert

Applying the card sorting technique, the field expert performed an intellectual-manual partition of the articles contained in the 17 clusters with the minimum size of three articles generated by the complete link cluster method.

1.2.1 The Partition

The partitioning was based on the conception of the field's division in specialties built up over the years, which served as a model for the partitioning of the articles here. This model basically involved five, possibly six specialties. The field expert had some hesitations which he overcame by referring to specific authors and their works. The field expert's clustering involved 10 clusters and the results are shown in Table 5-2.

Table 5-2: The Size-Frequency Distribution of Clusters Generated by the Field Expert

Size	Frequency	Size · Frequency	Percentage
10	1	10	16
9	2	18	29
8	1	8	13
7	1	7	11
6	1	6	10
5	2	10	16
3	1	3	5
1	1	1	2
Sum	10	63	100

Next, the field expert assigned labels to the clusters that he had generated to indicate the perceived research focus of clusters. These are listed in Table 5-3.

Table 5-3: The Field Expert's Labels

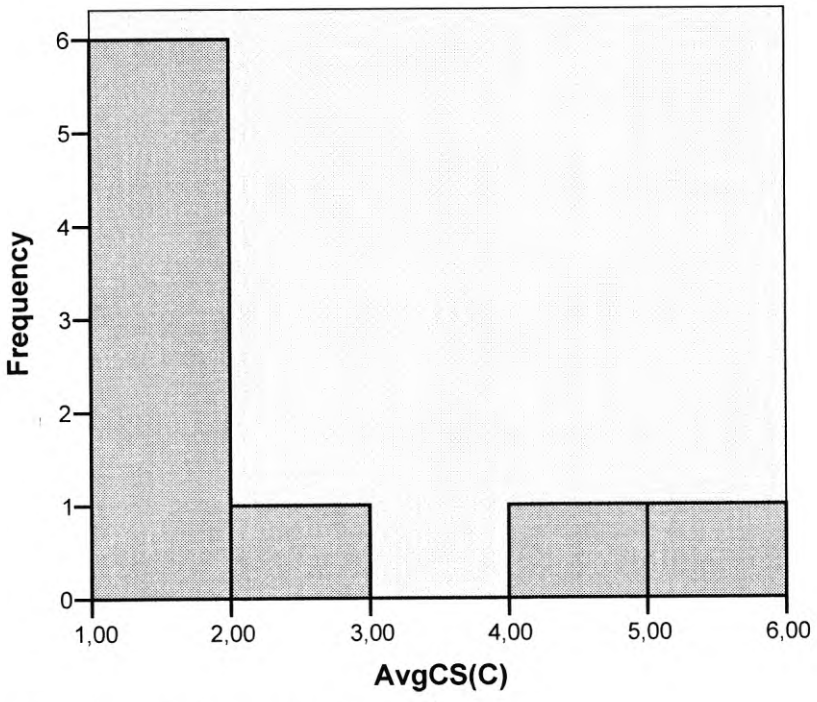
Cluster	Labels
1	Indicator development; journal impact factor; journal classification; measurement process
2	Mathematical distributions
3	Mapping
4	Collaboration
5	Webometrics
6	Science policy; science & technology; patents
7	Citation behavior
8	Case studies of particular fields
9	In memory of (V.V. Nalimov and B.C. Griffith)

Note: Cluster 10 (the singleton cluster) was not labeled.

1.2.2 Coherence and Separation

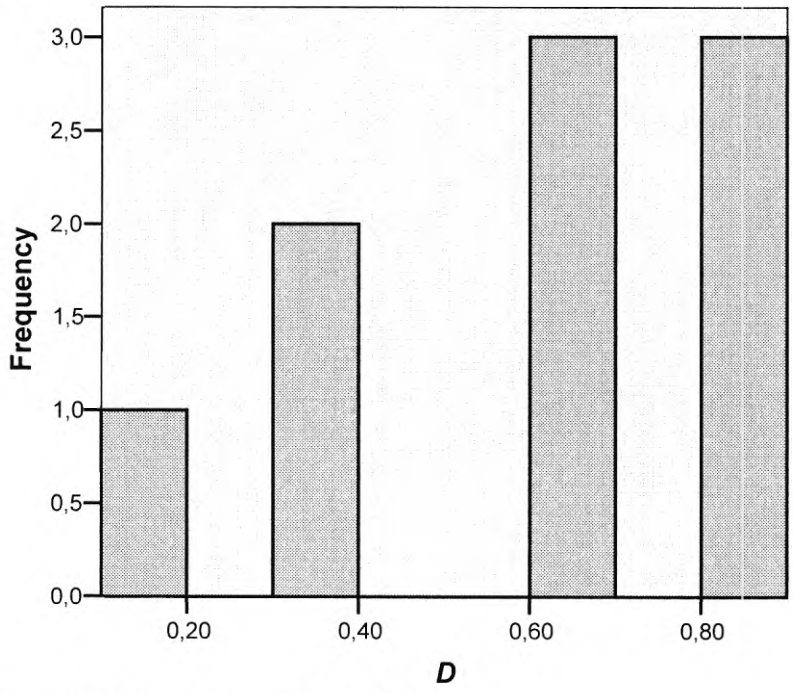
The coherence of the field expert's clusters in terms of bibliographic coupling units between articles in clusters was measured as the AvgCS(C) and *D* (see equation 2.2). The median AvgCS(C) was 1.90 and the median *D* was 0.61. The shapes of the resulting distributions are shown in Figures 5-3 and 5-4.

Figure 5-3: The Distribution of Coefficients of AvgCS(C) for Expert's Clusters in Case 1.



Note: Singleton clusters are not included.

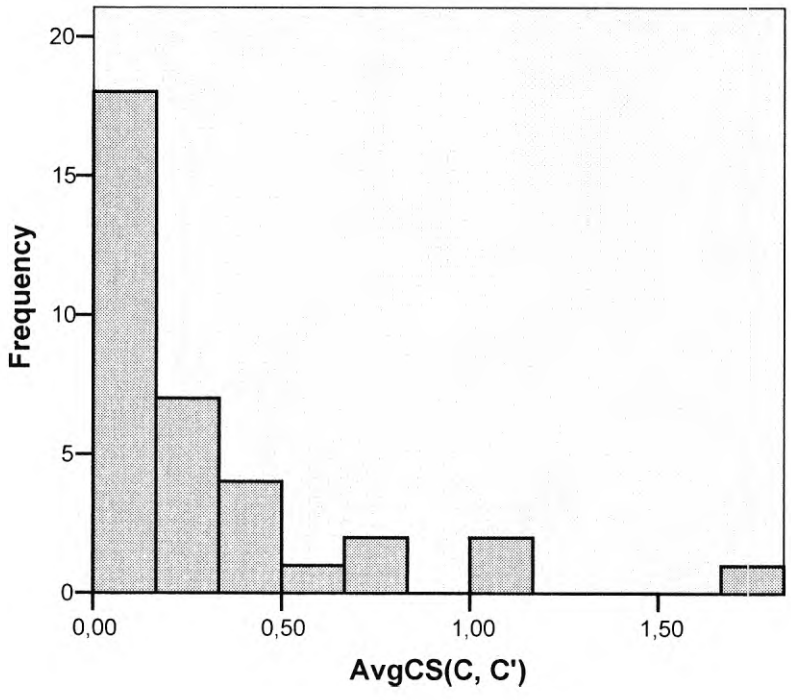
Figure 5-4: The Distribution of Coefficients of *D*



Note: Singleton clusters are not included

The separation between clusters was measured as the AvgCS(C, C'). Of all 2-combinations of clusters, 78 percent were coupled and no cluster was completely isolated. The median AvgCS(C, C') was 0.16 and the shape of the resulting distribution is shown in Figure 5-5.

Figure 5-5: The Distribution of Coefficients of AvgCS(C, C')



1.3 Analysis and Comparison of Partitions

The deviations between partitions generated by the field expert and by the complete link cluster method are presented with a point of departure in:

i. the coherence of clusters;

ii. the separation between clusters;

iii. the concentration of articles to clusters; and

iv. the qualitative assessment of cluster compositions.

For the sake of simplicity, the set of clusters with a size ≥ 3 generated by the complete link cluster method will be referred to as COMP and the set of clusters generated by the field experts partitions as EXP.

1.3.1 The Coherence of Clusters

With regard to AvgCS(C), both distributions are positively skewed, and the median value is higher for COMP. The median of *D* for EXP is 0.61, which should be compared with the default value of 1.0 for COMP. Conclusively, more coherent clusters were generated by the complete link cluster method.

1.3.2 The Separation Between Clusters

With regard to AvgCS(C, C'), both distributions are strongly positively skewed as most values are clustered at the lower intervals of the scale but the median for COMP is somewhat higher. The share of 2-combinations of clusters that are coupled is considerably lower for COMP. Conclusively, the general level of connectedness between clusters is lower for COMP, but links connecting clusters in COMP are slightly stronger.

1.3.3 The Concentration of Articles to Clusters

The two partitions differ much as the number of clusters in COMP was higher, 17 vs. 10 clusters in EXP. For COMP the value of Pratt's measure of concentration (see equation 4.4) is 0.13 and for EXP 0.39. Conclusively, articles are less concentrated to clusters in COMP.

1.3.4 The Qualitative Assessment of Cluster Compositions

A couple of clusters are near identical, and eight clusters in COMP constitute subsets of clusters in EXP. Two EXP clusters are completely split up by pairs of COMP clusters. The general pattern is that clusters in EXP are split up by two or three clusters in COMP. Conclusively, when compared, clusters in EXP are fragmented by clusters in COMP and two much deviating partitions are seen. For a detailed comparison of COMP and EXP with regard to the cluster compositions, see Appendix 3.

1.4 The Field Expert's Evaluation

The field expert performed a visual inspection of the set of clusters generated by the complete link cluster method, examining all articles in order to detect inconsistencies as to subject content in clusters. Any article deviating from the major subject theme of a cluster was regarded as misplaced and marked. In total, six articles (10 percent), all elements in two clusters, were marked as misplaced.

Comparing EXP with COMP, the field expert's view was that some deviations between partitions could be "renegotiated", foremost in terms of splitting up expert clusters.

1.5 Summary of Findings in Case 1

Below is a summary of the findings for this case.

i. More coherent clusters were generated by the complete link cluster method.

ii. Clusters generated by the complete link cluster method were less interconnected in terms of link density, though links between clusters were generally slightly stronger.

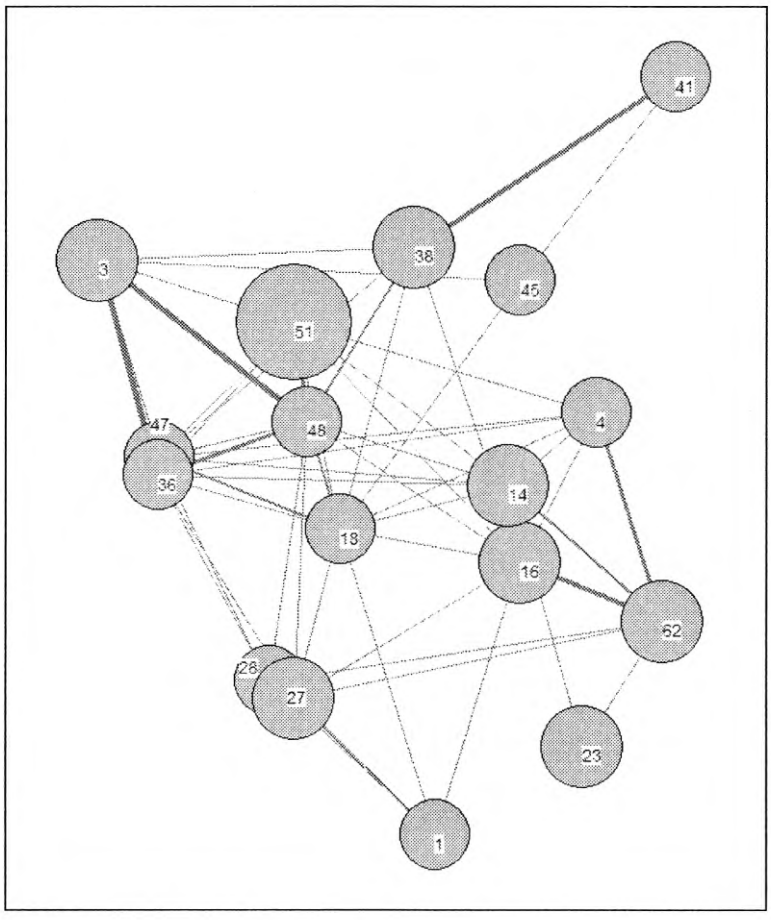
iii. The concentration of articles to clusters was lower for clusters generated by the complete link cluster method, which also was reflected by more and smaller clusters.

iv. Clusters generated by the complete link cluster method split up clusters generated by the field expert and presented a more fragmented picture of the analyzed research field.

v. Clusters generated by the complete link cluster method were generally relevant.

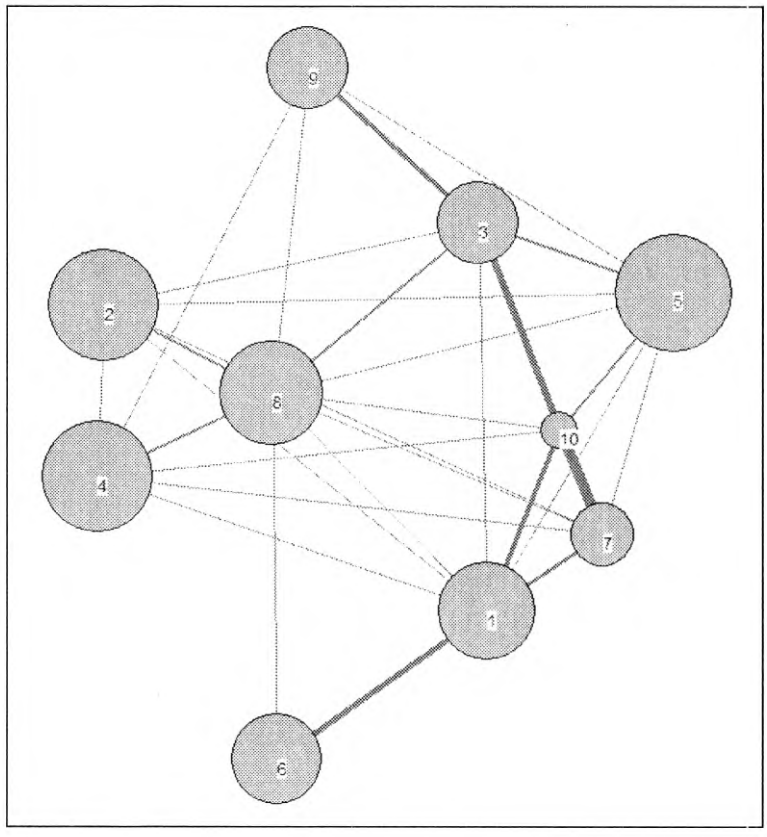
The basic features of the two deviating structures of clusters are depicted by MDS as graphs in Figures 5-6 and 5-7, where the distances between clusters are based on the AvgCS(C, C'). Here, circle sizes are proportional to cluster sizes and the width of connecting links to the strength of AvgCS(C, C'). These are not directly comparable between maps as there may be some scale difference. The sizes of a largest and a smallest cluster are given in the notes below the figures for guidance.

Figure 5-6: Graph of the 17 Clusters Generated by the Complete Link Cluster Method Visualized by MDS



Note: Kruskal's stress is 0.12. N for cluster 51 is 8 and N for cluster 48 is 3.

Figure 5-7: Graph of the 9 Clusters Generated by the Field Expert Visualized by MDS



Note: Kruskal's stress is 0.14. N for cluster 5 is 10 and N for cluster 10 is 1.

2. CASE 2: ORGANIC CHEMISTRY

2.1 Clusters Generated by the Complete Link Cluster Method

A total of 268 articles from the original population of 14,389 articles were merged, resulting in 95 clusters of which 17 were singleton clusters. Approximately 68 percent of all articles were grouped in clusters containing at least three articles (see Table 5-4). These were selected for further analysis. The bibliographic descriptions of articles in these clusters are presented in Appendix 4.

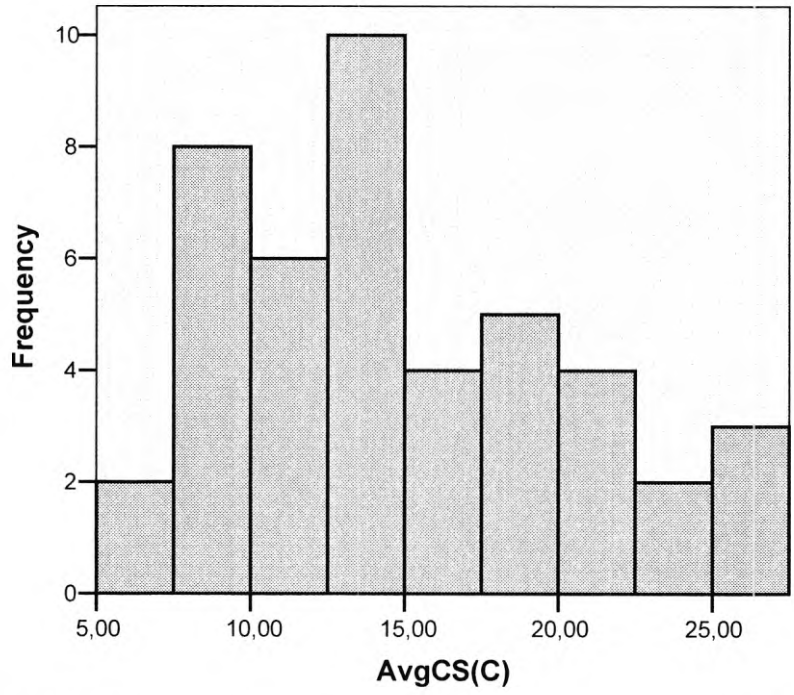
Table 5-4: The Size-frequency Distribution of Clusters Generated by the Complete Link Cluster Method

Size	Frequency	Size · frequency	Percentage
8	1	8	3
7	2	14	5
6	3	18	7
5	11	55	21
4	7	28	10
3	20	60	22
2	34	68	25
1	17	17	6
Sum	95	268	100

2.1.1 Coherence and Separation

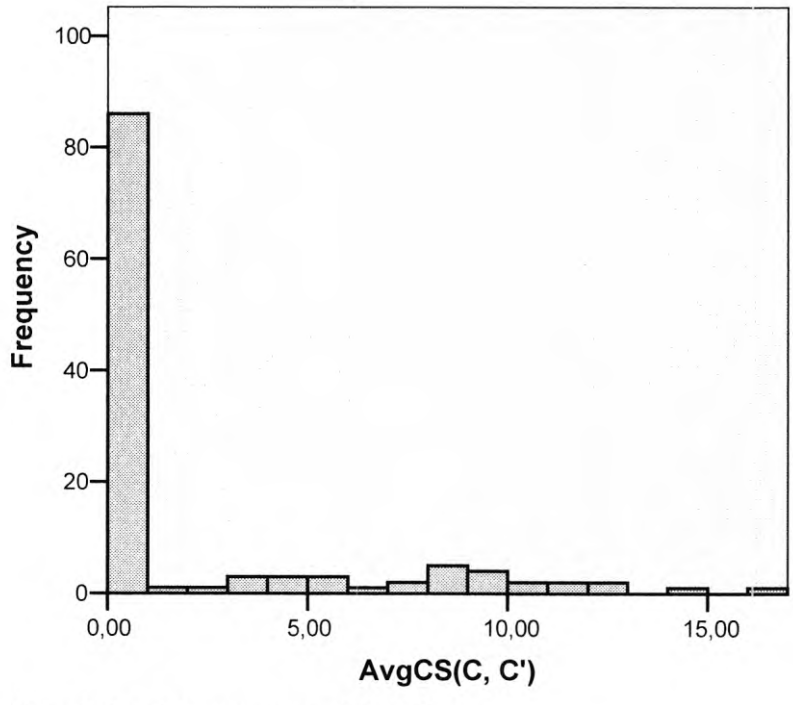
For 44 selected clusters containing 183 articles, the AvgCS(C) (see equation 4.2) was measured. The median AvgCS(C) was 13.94 and the shape of the resulting distribution is shown in Figure 5-8.

Figure 5-8: The Distribution of Coefficients of AvgCS(C)



Next, the separation between clusters was measured as the AvgCS(C, C') (see equation 4.3). Of all 2-combinations of clusters, 12 percent were coupled and five clusters were completely isolated. The median AvgCS(C, C') was 0.24 and the shape of the resulting distribution is shown in Figure 5-9.

Figure 5-9: The Distribution of Coefficients of AvgCS(C, C')



Note: Isolated clusters are not included.

2.2 Core Documents - a Microanalysis

As the size and character of the field allows for the identification of core documents, a microanalysis of the application of the proposed method's applicability for core document mapping was pursued. In the network of shared references, a total of 49 source articles fulfilled the requirements for core documents as defined (see Sub-section 2.1 in Chapter 2). In this set, all but two articles were bibliographically coupled with another core document. Computing bibliographic couplings between articles in this set, the exclusion of pairs sharing less than two references resulted in a total of 186 coupled pairs, 2,333 bibliographic couplings and 46 articles. The range of NCS was 0.88 (i.e. 0.024 – 0.90) and the median coefficient was 0.30. A matrix, containing $\binom{46}{2} = 1{,}035$ elements based on this set was computed, which could be considered a rather dense subgraph where the density D (see equation 2.2) was 0.17.[43] This matrix was then used as input for complete link clustering and MDS. It showed up as seven clearly demarcated clusters on the MDS map in full agreement with clusters generated by the complete link cluster method (see Figure 5-14 in Sub-section 2.5 in this Section).

2.3 Clusters Generated by the Field Expert

Applying the card sorting technique, the field expert performed an intellectual partition of the articles contained in the 44 clusters with the minimum size of three articles, generated by the complete link cluster method.

2.3.1 The Partition

The field expert in this case used a principle of division based on four major categories:

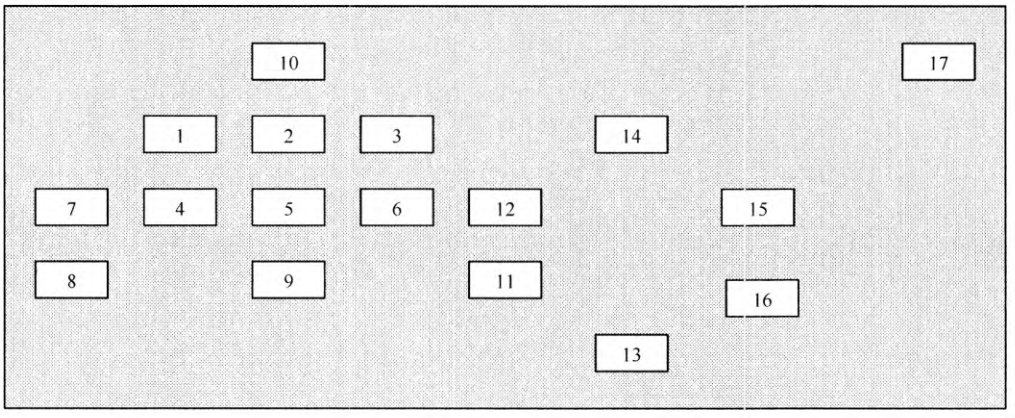
i. properties;
ii. synthesis;
iii. understanding of mechanism; and
iv. method.

These four categories were in turn subdivided.

The rectangular table used for the card sorting was applied in a way that a spatial representation was accomplished (see Figure 5-10).

[43] This could be compared with the share of 2 percent of 2-combinations of articles that were coupled in the original matrix.

Figure 5-10: The Configuration of Piles on the Card Sorting Table



The configuration of 17 piles of cards of bibliographic representations of articles were placed in a way that piles (clusters) with a similar or connecting research foci were located in each others vicinity and each pile was assigned a label indicating the perceived research focus (see Table 5-5).

Table 5-5: The Field Expert's Labels

Cluster	Labels
1	Synthesis product
2	Synthesis reaction
3	Catalysis
4	Stereo selective synthesis-product
5	Stereo selective synthesis-reactions
6	Stereo selective synthesis-catalysis
7	Total synthesis
8	Total synthesis & medicinal chemistry
9	Stereo selective synthesis-racemization
10	Synthesis evaluation
11	Peptide synthesis
12	DNA templated organic synthesis
13	Stereo selective reaction mechanism
14	Reaction mechanism
15	DNA properties
16	Peptide structure
17	Nano

All articles were contained in clusters with a size of at least three articles, with the exception of three singleton clusters. The distribution of articles over clusters is shown in Table 5-6.

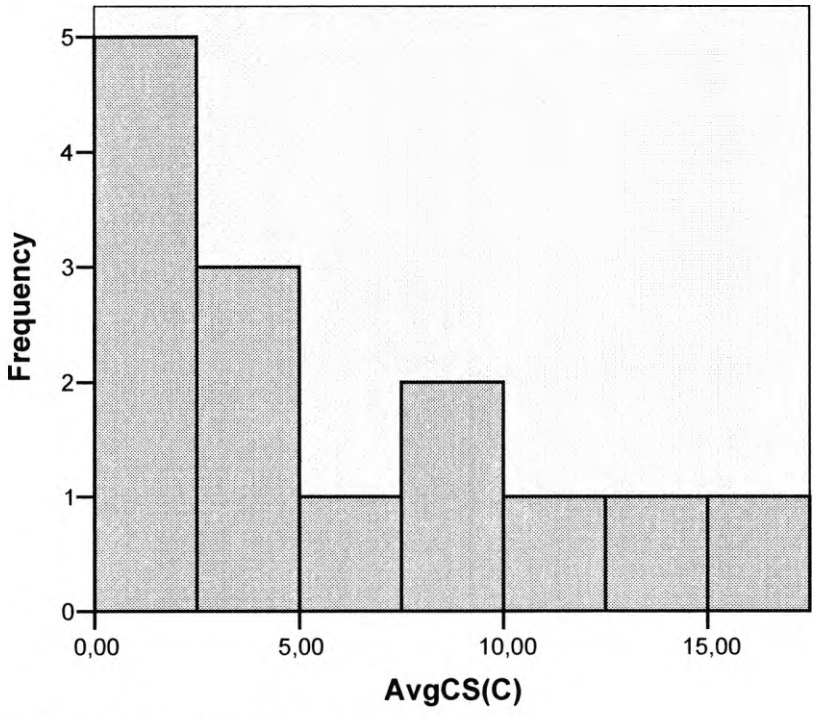
Table 5-6: The Size-frequency Distribution of Clusters Generated by the Field Expert's Clustering

Size	Frequency	Size · Frequency	Percentage
43	1	43	23
29	1	29	16
25	1	25	14
14	2	28	15
13	1	13	7
12	1	12	7
10	1	10	5
5	1	5	3
3	5	15	8
1	3	3	2
Sum	17	183	100

2.3.2 Coherence and Separation

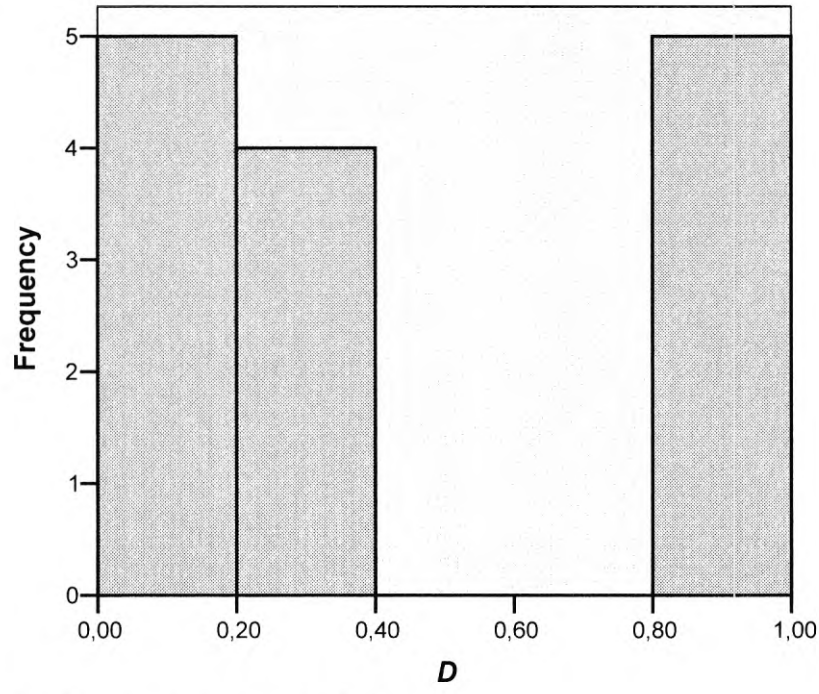
The coherence of the field expert's clusters in terms of bibliographic coupling units between articles in clusters was measured as the AvgCS(C) and *D* (see equation 2.2). The median AvgCS(C) was 3.91 and the median *D* 0.30. The shapes of the resulting distributions are shown in Figures 5-11 and 5-12.

Figure 5-11: The Distribution of Coefficients of AvgCS(C)



Note: Singleton clusters are not included.

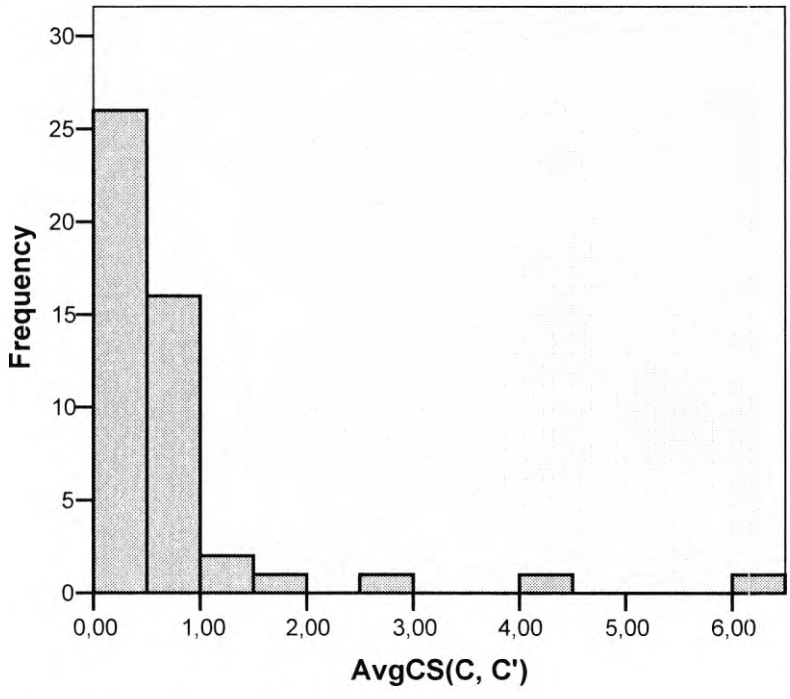
Figure 5-12: The Distribution of Coefficients of *D*



Note: Singleton clusters are not included.

The separation between clusters was measured as the AvgCS(C, C'). Of all 2-combinations of clusters, 35 percent were coupled and four clusters were found completely isolated. The median AvgCS(C, C') was 0.40 and the shape of the resulting distribution is shown in Figure 5-13.

Figure 5-13: The Distribution of Coefficients of AvgCS(C, C')



Note: Isolated clusters are not included.

2.4 Analysis and Comparison of Partitions

The deviations between partitions generated by the field expert and by the complete link cluster method are presented with a point of departure in:

i. the coherence of clusters;

ii. the separation between clusters;

iii. the concentration of articles to clusters; and

iv. the qualitative assessment of cluster compositions.

For the sake of simplicity, the set of clusters with a size ≥ 3 generated by the complete link cluster method will be referred to as COMP and the set of clusters generated by the field experts partitions as EXP.

2.4.1 The Coherence of Clusters

With regard to AvgCS(C), the distribution of COMP is more normally distributed, whereas the distribution of EXP is slightly positively skewed. The two distributions are deviating, as the median of COMP is more than three times the median of EXP. The median of *D* for EXP is 0.30 which should be compared with the default value of 1.0 for COMP. Conclusively, clusters in COMP are generally much more coherent.

2.4.2 The Separation between Clusters

Both COMP and EXP have isolated clusters, five and four respectively. With regard to AvgCS(C, C'), both distributions are strongly positively skewed with extremes. For COMP, the extremes are more frequent and clustered at a higher level of the scale, indicating relatively strong associations between several clusters. Also, the median AvgCS(C, C') for COMP is lower. The share of 2-combinations of clusters that are coupled is considerably lower for COMP too. Conclusively, the general level of connectedness between clusters in COMP is lower, however, a considerable share of relatively strong links (extremes) between more than 20 clusters was found in the case of COMP (all extreme links can are reflected as clusters of clusters in the map in Figure 5-15 shown under Sub-section 2.5 in this chapter).

2.4.3 The Concentration of Articles to Clusters

The two partitions differ much as the number of clusters in Comp was higher, 44 vs. 17 clusters in EXP. This involves a difference between the two partitions with regard to the concentration of articles to clusters. The value of Pratt's measure of concentration (see equation 4.4) is 0.17 for COMP and 0.57 for EXP. Conclusively, articles were more concentrated to clusters in the case of EXP.

2.4.4 The Qualitative Assessment of Cluster Compositions

The most striking deviation is the difference in cluster sizes which is pronounced by three large expert clusters with a size between 43 and 25. Over these three clusters, approximately half of all articles are dispersed and the number of intersections with COMP is 53. Several (14) COMP clusters are completely contained within an EXP cluster, and approximately 30 percent of all articles belong to this category. Therefore, several of COMP's clusters constitute fractions of EXP's clusters. It can be concluded that two very deviating classification systems have been operating. For a detailed comparison of COMP and EXP with regard to the cluster compositions, see Appendix 5.

2.5 The Field Expert's Evaluation

When evaluating the partition generated by the complete link cluster method, the field expert concluded that "there exists several possible common denominators" for a single cluster solution, and these were often hard to decide. In fact, the proposed method supplied the expert with a new set of principles of division which were not easy to anticipate. Some of the classes brought about by the complete link clustering could clearly be of interest in relation to several research questions at issue. A number of clusters were based on the association between articles with a common focus on specific methods whereas others seemed to be based on the association between articles with a common focus on chemical compounds or classes of compounds.

Regarding the evaluation of clusters, all 44 clusters were inspected article by article in order to detect inconsistencies as to subject content in clusters where any article deviating from the major subject theme of a cluster was regarded as misplaced and marked. In total, four articles in four different clusters were regarded as deviating from the cluster's research focus.

Concerning the map of core documents, the expert performed an evaluation of both the clusters and the configuration of the map (see Figure 5-14). One article (13439) was considered unclassifiable, as the expert was not familiar with the subject presented by this article. Except for this article, all other articles were considered correctly located. In the configuration in the MDS map, two dimensions could be discerned:

i. **Top-bottom**

"Non-stereo selective reactions" to "stereo selective reactions".

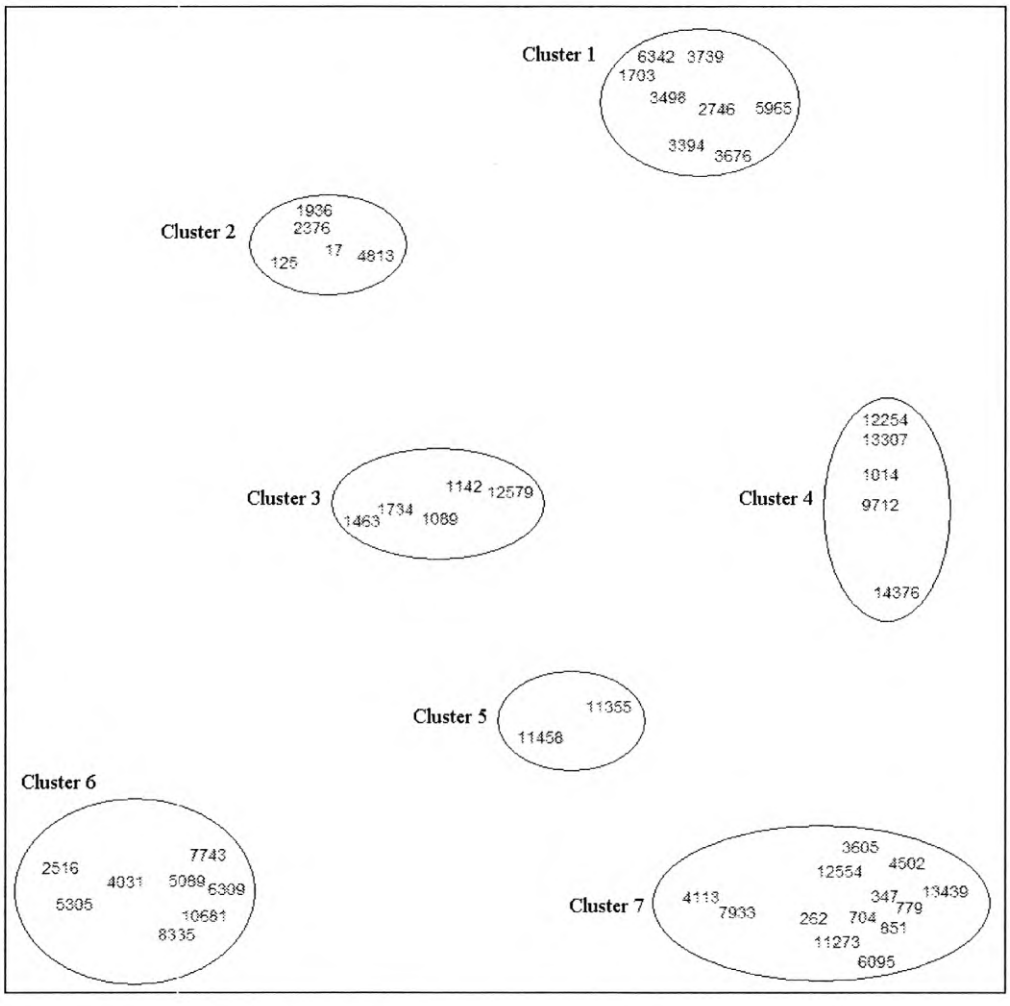
ii. **Left-right**

"Metal-catalysis" to "other chemistry".

The configuration of clusters in the MDS map was considered intelligible, though cluster 1 should be more distant to the other clusters as its research

theme (DNA) was divergent in relation to the research themes of all the other clusters in the map. The bibliographic descriptions of evaluated core documents are presented in Appendix 6.

Figure 5-14: Map of the Associations between 46 Core Documents Generated by MDS and the Complete Link Cluster Method



Note: Kruskal's stress is 0.03.

2.6 Summary of Findings in Case 2

Below is a summary of the findings for this case.

i. More coherent clusters were generated by the complete link cluster method.

ii. Clusters generated by the complete link cluster method were less interconnected in terms of link density, and generally, links were also weaker. However, several clusters generated by the complete link cluster method were associated by strong links (extremes).

iii. The concentration of articles to clusters was weaker for clusters generated by the complete link cluster method, which also was reflected by more and smaller clusters.

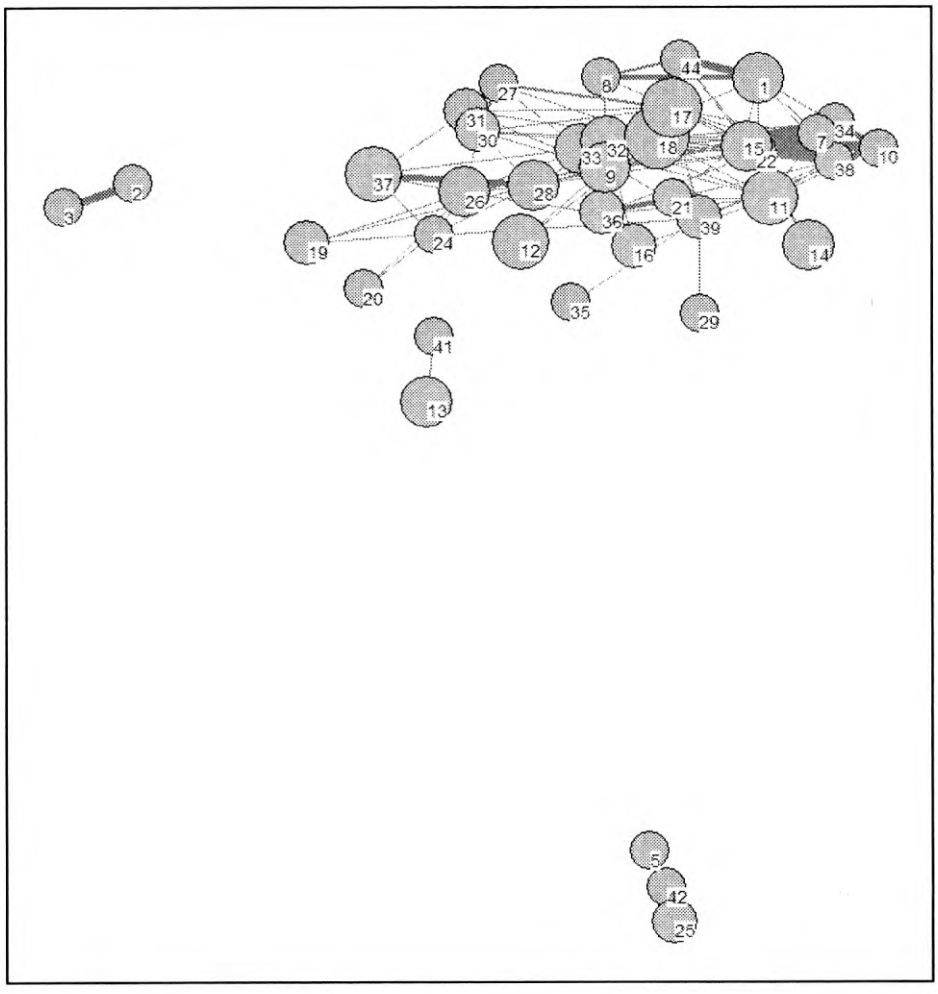
iv. Clusters generated by the complete link cluster method presented a more fragmented picture of the analyzed research field and there is little agreement between the two partitions.

v. Clusters generated by the complete link cluster method were mostly completely relevant.

vi. The clustering and mapping of core documents identified and depicted highly relevant clusters and the configuration of clusters when mapped was basically relevant.

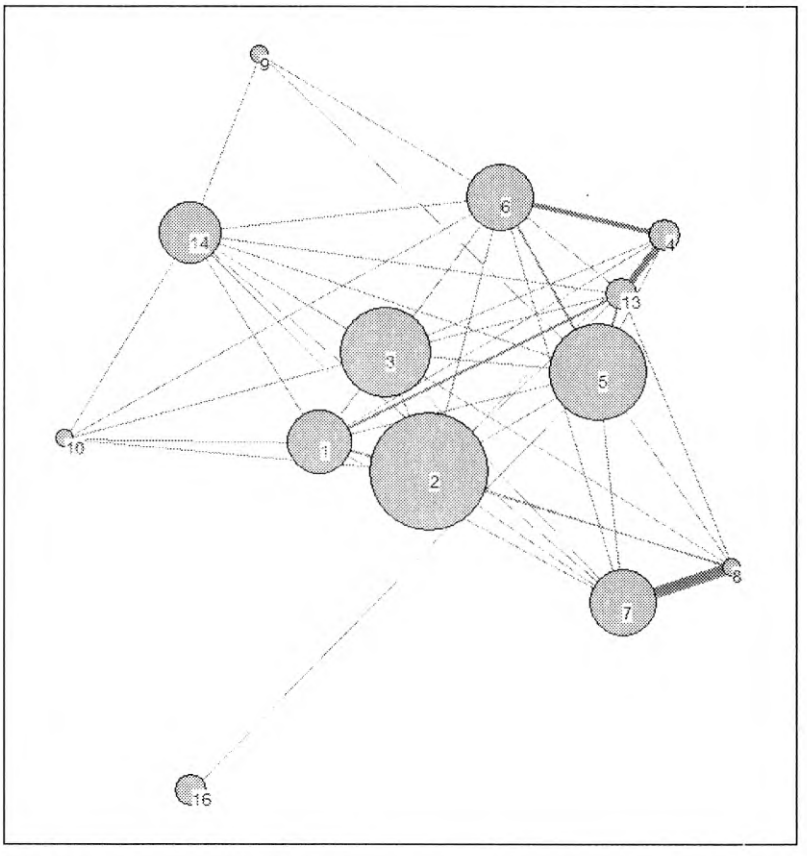
The basic features of the two deviating structures of clusters are depicted by MDS as graphs in Figures 5-15 and 5-16, where the distances between clusters are based on the AvgCS(C, C'). Here, circle sizes are proportional to cluster sizes and the width of connecting links to the strength of AvgCS(C, C'). These are not directly comparable between maps as there may be some scale difference. The sizes of a largest and a smallest cluster are given in the notes below the figures for guidance.

Figure 5-15: Graphs of the 39 Clusters Generated by the Complete Link Cluster Method Visualized by MDS



Note: Clusters 4, 6, 23, 40 and 43 are not mapped as they are isolated. N for cluster 18 is 8 and N for cluster 35 is 3. Kruskal's stress is 0.04.

Figure 5-16: Graph of the 13 Clusters Generated by the Field Expert Visualized by MDS



Note: Clusters 11, 12, 15 and 17 are not mapped as they are isolated. N for cluster 2 is 43 and N for cluster 10 is 1. Kruskal's stress is 0.13.

3. CASE 3: PURE & APPLIED MATHEMATICS

3.1 Clusters Generated by the Complete Link Cluster Method

A total of 579 articles from the original population of 879 articles were merged, resulting in 420 clusters of which 311 were singleton clusters. Approximately 22 percent of all articles were grouped in clusters containing at least three articles (see Table 5-7). These were selected for further analysis. The bibliographic descriptions of articles in these clusters are presented in Appendix 7.

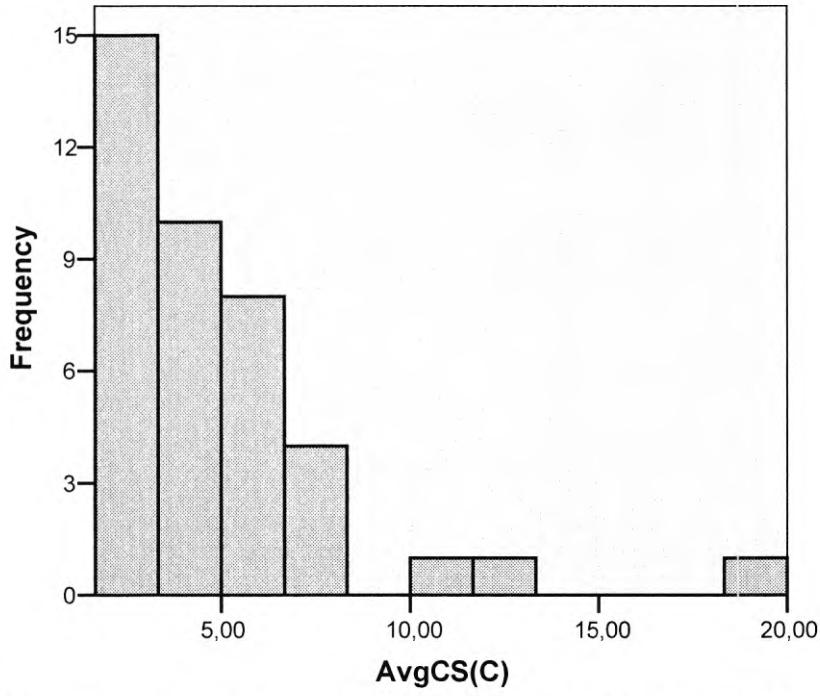
Table 5-7: The Size-Frequency Distribution of Clusters Generated by the Complete Link Cluster Method

Size	Frequency	Size · frequency	Percentage
6	1	6	1
5	1	5	1
4	5	20	3
3	33	99	17
2	69	138	24
1	311	311	54
Sum	420	579	100

3.1.1 Coherence and Separation

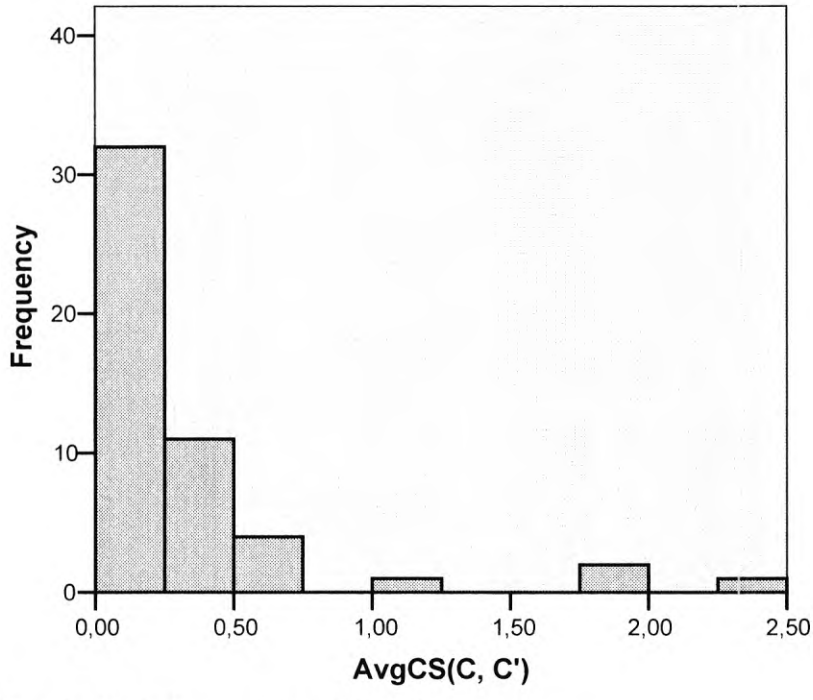
For 40 selected clusters containing 130 articles, the AvgCS(C) (see equation 4.2) was calculated. The median was 4.17 and the shape of the resulting distribution is summarized in Figure 5-17.

Figure 5-17: The Distribution of Coefficients of AvgCS(C)



Next, the separation between clusters was calculated as the AvgCS(C, C') (see equation 4.3). Of all 2-combinations of clusters, seven percent were coupled and five clusters were completely isolated. The median was 0.17 and the shape of the resulting distribution is shown in Figure 5-18.

Figure 5-18: The Distribution of Coefficients of AvgCS(C, C')



Note: Isolated clusters are not included.

3.2 Clusters Generated by the Field Expert

Applying a card sorting technique, the field expert performed an intellectual-manual partition of the articles contained in the 40 clusters with the minimum size of three articles generated by the complete link cluster method.

3.2.1 The Partition

When asked to describe the mode of procedure used to perform the clustering, the field expert's reply was that he referred to the classification or intellectual map of sub-fields obtained during the years of mathematical studies and that structures of curricula lay a basis for the apprehension of the field's dividing up in specialties. According to the field expert, there was a difficulty of deciding a point of departure for the intellectual classification. A principle of partition could have two different starting points: (i) the "event" and (ii) the "scene". Here, (i) should aim at what is actually performed, and (ii) at the context of a particular performance and the selection of either point of departure was done in an arbitrary way. The field expert emphasized the subjectiveness of the intellectual clustering and pointed out that a more fine graded partition as well as additional mergings of groups could be equally valid from his point of view. In concordance with this, when labeling clusters, the field expert provided on several occasions two classifications for a cluster where the first classification specified a sub-field to the second. The classification of 36 clusters by assigned labels is given in Table 5-8 and the distribution of articles over clusters in Table 5-9.

Table 5-8: The Field Expert's Labels

Cluster	First Classification	Second Classification
1	Symplectic manifolds	differential typology
2	Convexity	-
3	Homotopy theory	algebraic typology
3	Homotopy theory	-
4	Probability theory	differential typology
5	Invariant theory	differential typology
6	Kahler manifolds	differential typology
7	a fourth order PDE	dynamical systems
8	C*-algebras	Banach-algebras
9	Complex manifolds	differential typology
10	Riemannian manifolds	differential typology
11	Differentials and foliations	differential typology
12	Quadratic mappings	-
13	Category theory	abstract algebra
14	Foliation manifolds	differential typology
15	Homology - Cohomology	algebraic topology
16	Projective spaces-surfaces-	algebraic geometry
16	Projective spaces-surfaces-	algebraic geometry
17	Grassman manifolds - Flog	differential typology
18	Group theory	-
19	Elliptic curves	algebraic geometry
20	P-adic theory, dyadic	-
21	Measure theory	-
22	KdV equations	partial differential equations
23	Theory for algebras; Lie, Kac-	-
24	Differential geometry	-
25	Klein-Gordon equations,	-
26	Elliptic differential equations	-
27	Variation analysis	-
28	Homogenous spaces	differential typology
29	Holomorphic mappings	-
30	Operator-algebras	-
31	Representation theory	-
32	Dynamic systems	-
33	Sheaves	differential typology
34	Banach space	-
35	Fields with valuations	algebraic geometry
36	Parabolic differential equations	-

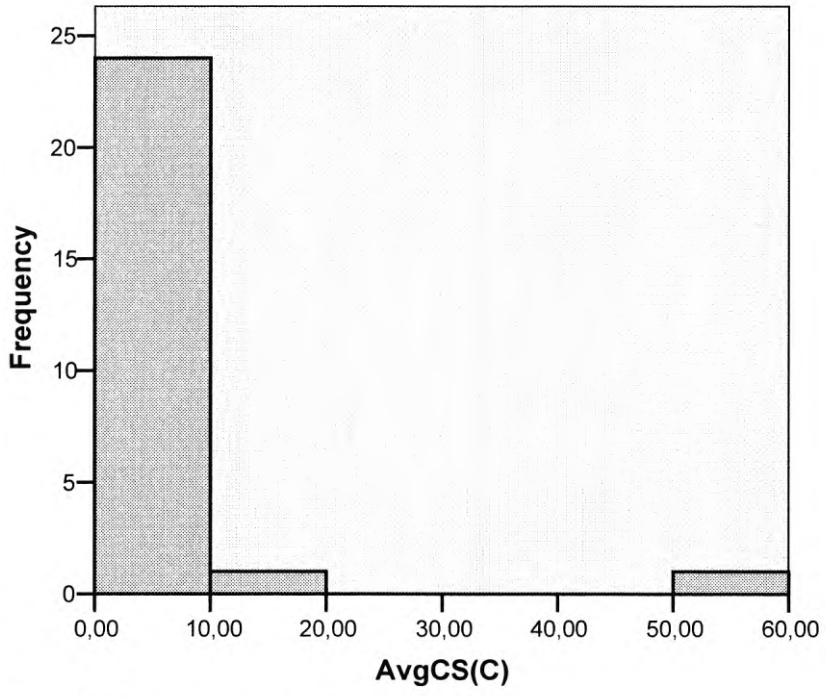
Table 5-9: The Size-Frequency Distribution of Clusters Generated by the Field Expert

Size	Frequency	Size · Frequency	Percentage
16	1	16	13
7	6	42	33
6	2	12	9
4	4	16	13
3	5	15	12
2	8	16	13
1	10	10	8
Sum	36	127	100

3.2.2 Coherence and Separation

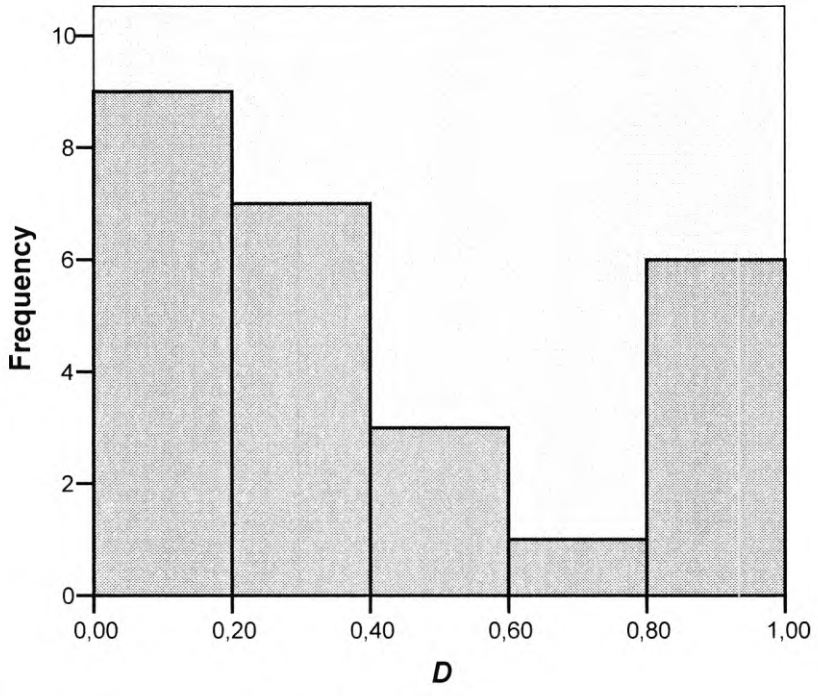
The coherence of the field expert's clusters in terms of bibliographic coupling units between articles in clusters was measured as the AvgCS(C) and *D* (see equation 2.2). The median AvgCS(C) was 1.65 and the median *D* was 0.33. The shapes of the resulting distributions are shown in Figures 5-19 and 5-20.

Figure 5-19: The Distribution of Coefficients of AvgCS(C)



Note: Singleton clusters are not included.

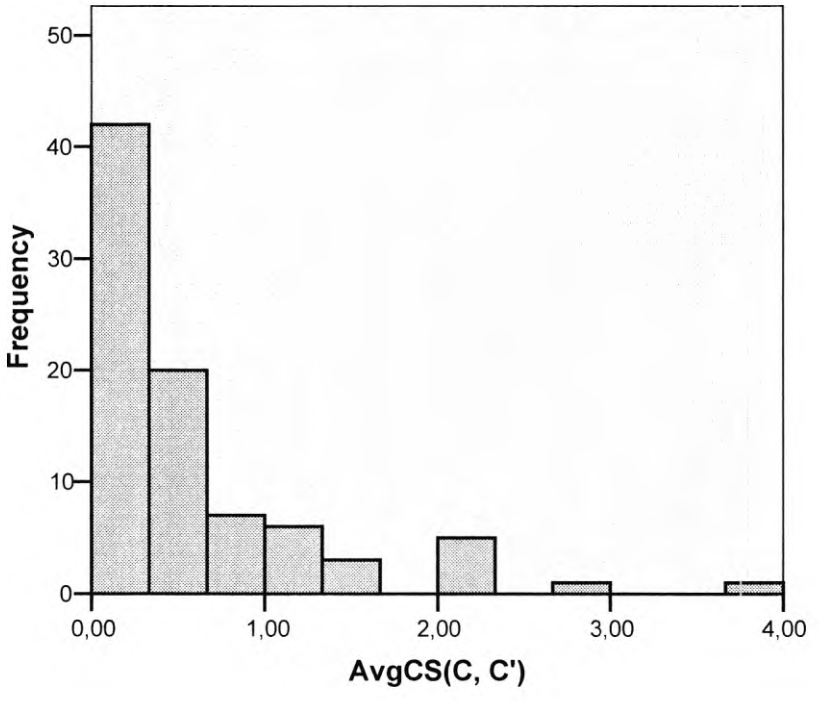
Figure 5-20: The Distribution of Coefficients of *D*



Note: Singleton clusters are not included.

The separation between clusters was measured as the AvgCS(C, C'). Of all 2-combinations of clusters, 13 percent were coupled and no cluster was completely isolated. The median AvgCS(C, C') was 0.36 and the shape of the resulting distribution is shown in Figure 5-21.

Figure 5-21: The Distribution of Coefficients of AvgCS(C, C')



3.3 Analysis and Comparison of Partitions

The deviation between partitions generated by the field expert and by the complete link cluster method are presented with a point of departure in:

i. the coherence of clusters;

ii. the separation between clusters;

iii. the concentration of articles to clusters; and

iv. the qualitative assessment of cluster compositions.

For the sake of simplicity, the set of clusters with a size ≥ 3 generated by the complete link cluster method will be referred to as COMP and the set of clusters generated by the field experts partitions as EXP.

3.3.1 The Coherence of Clusters

With regard to AvgCS(C), both distributions are positively skewed with some extremes. For EXP, the extremes refer to a couple of clusters with the size of two, which have similar subject foci and very similar reference lists. The median value of AvgCS(C) was considerably higher for COMP. With regard to *D*, the median for EXP was 0.33, which should be compared with the default value of 1.0 for COMP. Conclusively, more coherent clusters were generated by the complete link cluster method.

3.3.2 The Separation between Clusters

With regard to AvgCS(C, C'), both distributions are positively skewed with a few extremes each, and the median is lower for COMP. In addition, five clusters in COMP were isolated. The share of 2-combinations of clusters that are coupled is lower for COMP, though for both distributions holds that the general level of interconnectedness between clusters is low. Conclusively, the general level of interconnectedness between clusters is lower for COMP, and links connecting clusters in COMP are generally weaker too.

3.3.3 The Concentration of Articles to Clusters

The two partitions differ with regard to the dispersion of articles over clusters. The articles contained in 40 clusters in COMP, are dispersed over 26 clusters in EXP and an additional ten articles are singleton clusters. The value of Pratt's measure of concentration (see equation 4.4) is 0.07 for COMP and 0.42 for EXP. Conclusively, articles are less concentrated to clusters in COMP.

3.3.4 The Qualitative Assessment of Cluster Compositions

Three clusters, all with the size of two, are identical, and eight clusters in COMP constitute subsets of clusters in EXP. Two EXP clusters are completely split up by pairs of COMP clusters. In all, clusters in COMP provide a more

fragmented depiction of the field and two much deviating partitions are seen. For a detailed comparison of COMP and EXP with regard to the cluster compositions, see Appendix 8.

3.4 The Field Expert's Evaluation

A visual inspection of the clusters generated by the complete link cluster method was performed by the field expert and 40 clusters were inspected article by article in order to detect inconsistencies as to subject content in clusters and any article deviating from the major subject theme of a cluster was regarded as misplaced and marked. A total of three articles were regarded as unclassifiable in the sense that titles and additional bibliographic information added insufficient information regarding topics. In total, 17 articles (13 percent) were marked as misplaced.

3.5 Summary of Findings in Case 3

Below is a summary of the findings for this case.

i. More coherent clusters were generated by the complete link cluster method;

ii. Clusters generated by the complete link cluster method were less interconnected in terms of link density, and links between clusters were generally weaker.

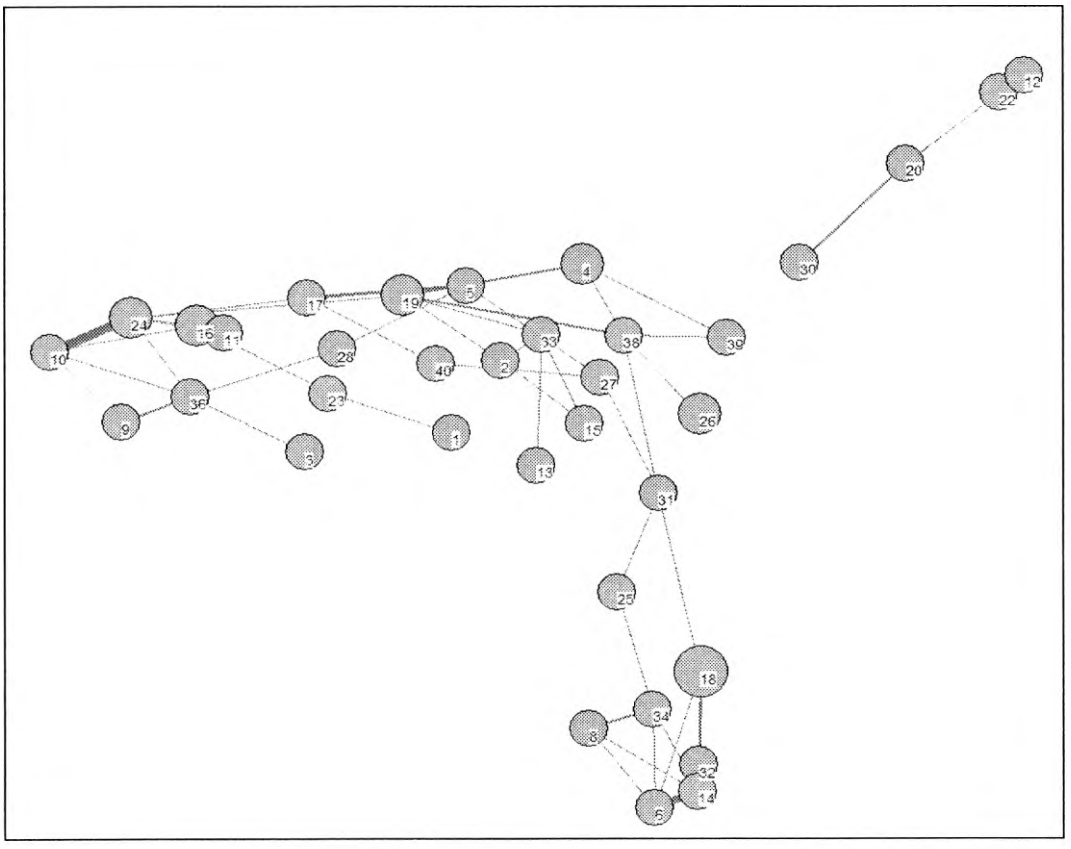
iii. The concentration of articles to clusters was lower for clusters generated by the complete link cluster method. The dispersion of articles differed considerably between the two partitions, foremost with regard to the evenness of the distribution of cluster sizes where clusters generated by the complete link cluster method showed a lower variation.

iv. The cluster structure generated by the complete link cluster method depicted the analyzed research field as more fragmented.

v. Clusters generated by the complete link cluster method were generally relevant.

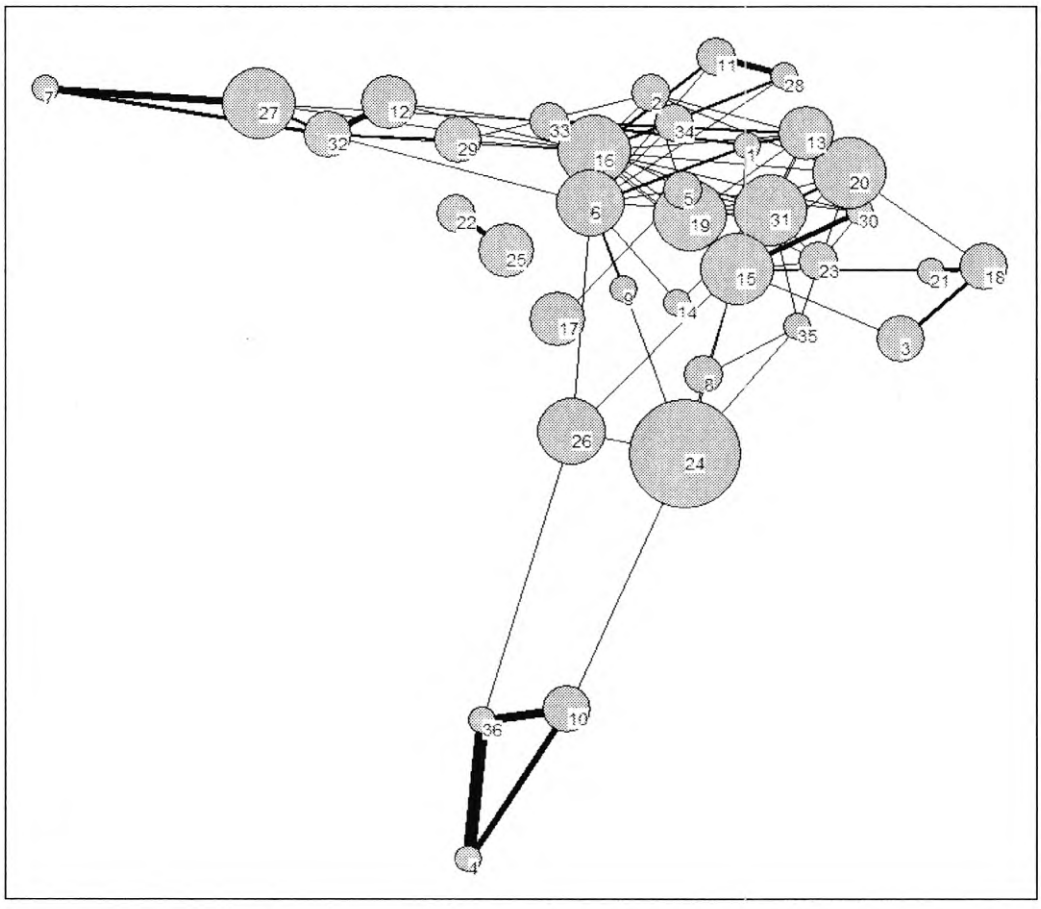
The basic features of the two deviating structures of clusters are depicted by MDS as graphs in Figures 5-22 and 5-23, where the distances between clusters are based on the AvgCS(C, C'). Here, circle sizes are proportional to cluster sizes and the width of connecting links to the strength of AvgCS(C, C'). These are not directly comparable between maps as there may be some scale difference. The sizes of a largest and a smallest cluster are given in the notes below the figures for guidance.

Figure 5-22: Graphs of the 35 Clusters Generated by the Complete Link Cluster Method Visualized by MDS



Note: Clusters 7, 21, 29, 35 and 37 are not mapped as they are isolated. N for cluster 18 is 6 and N for cluster 34 is 3. Kruskal's stress is 0.03

Figure 5-23: Graph of the 36 Clusters Generated by the Field Expert Visualized by MDS



Note: Kruskal's stress is 0.07. N for cluster 24 is 16 and N for cluster 8 is 1.

4. CASE 4: CORE DOCUMENTS

The findings of this case is presented in a different manner from the previous four cases in view of its complexity and different research questions. For simplicity purpose, the following notations will be used when referring to the different levels at which clusters were generated:

i. C1 denotes the level at which clusters were generated by the first clustering. Applied measure of document similarity was NCS and the cluster method was the complete link cluster method. From here and onwards, clusters generated at the C1 level are referred to as "C1" with an added cluster identification number (e.g. C1/32) when a particular cluster is referred to. When the complete identity of a C1 cluster is addressed, all levels are noted, e.g. C3/3/C2/87/C1/1931.

ii. C2 denotes the level at which clusters were generated by the second clustering. Applied measure of cluster similarity was AvgCS(C, C') and the cluster method was again the complete link cluster method. From here and onwards, clusters generated at the C2 level are referred to as "C2" with an added cluster identification number when a particular cluster is referred to.

iii. C3 denotes the third and last level of clustering. Applied measure of cluster similarity was AvgCS(C, C') and the method of partition was the between groups average cluster method. From here and onwards, clusters generated at the C3 level are referred to as "C3" with an added cluster identification number when a particular cluster is referred to.

The findings are presented in the following order:

1. Results from the first clustering (C1-level)

2. Results from the iterated clustering (C2-level)

3. Results from the reiterated clustering (C3-level)

4. Field experts' evaluations of four cases of reiterated clustering

5. Results from the expansion of C1-clusters

6. Summary of findings

Though the presentation of findings should aim at an order that is in line with the sequence of pursued experiments, for practical reasons, results from the clustering on the different levels of fusion are presented in one sequence in order to facilitate the comprehension of changes. However, point 5 preceded points 2 and 5 empirically.

The fusions of clusters at the different levels are illustrated by examples. With the starting point in a C1-cluster, the stepwise merging of clusters to a final

C3-cluster is described. The selection of the C3-cluster was based on its approachability on a layman level in terms of comprehensible core document titles.[44] With regard to point 5, an example of an expanded C1-cluster is also given.

4.1 The First Fusion Level - C1 Clusters

4.1.1 Clusters and Cluster Sizes

In order to assess the association through bibliographic coupling between core documents, the NCS between all 6,060 core documents of the original population was computed (see Sub-section 3.2.5 in Chapter 4). Links with a NCS lower than 0.25 were filtered out and 5,771 articles were clustered by the complete link cluster method. A total of 1,761 clusters were generated of which 228 were singleton clusters. In all, 5,543 core documents were merged to 1,533 clusters varying in size between 2 and 22. 1000 clusters had a size ≥ 3 and contained in total 4,477 core documents (see Table 5-10). These 1000 clusters were selected for further analysis and fusion.

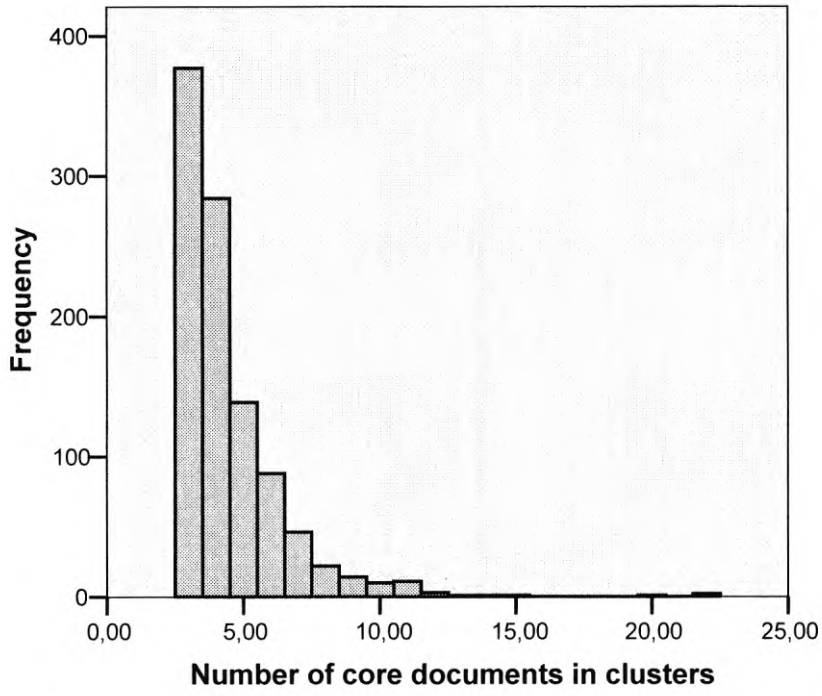
Table 5-10: The Size-Frequency Distribution of C1-Clusters

Size	Frequency	Size · Frequency	Percentage
22	2	44	0.76
20	1	20	0.35
15	1	15	0.26
14	1	14	0.24
13	1	13	0.23
12	3	36	0.62
11	11	121	2.10
10	10	100	1.73
9	14	126	2.18
8	22	176	3.05
7	46	322	5.58
6	88	528	9.15
5	139	695	12.04
4	284	1136	19.68
3	377	1131	19.60
2	533	1066	18.47
1	228	228	3.95
Sum	1761	5771	100.0

The shape of the distribution of core documents over C1-clusters where the size is ≥ 3 is summarized in Figure 5-24.

[44] The author of this thesis has a background in health sciences, and associations between articles and clusters could be interpreted, at least in a superficial way, as the general topic of the example cluster is about infectious diseases.

Figure 5-24: The Distribution of Core Documents over 1000 C1-Clusters

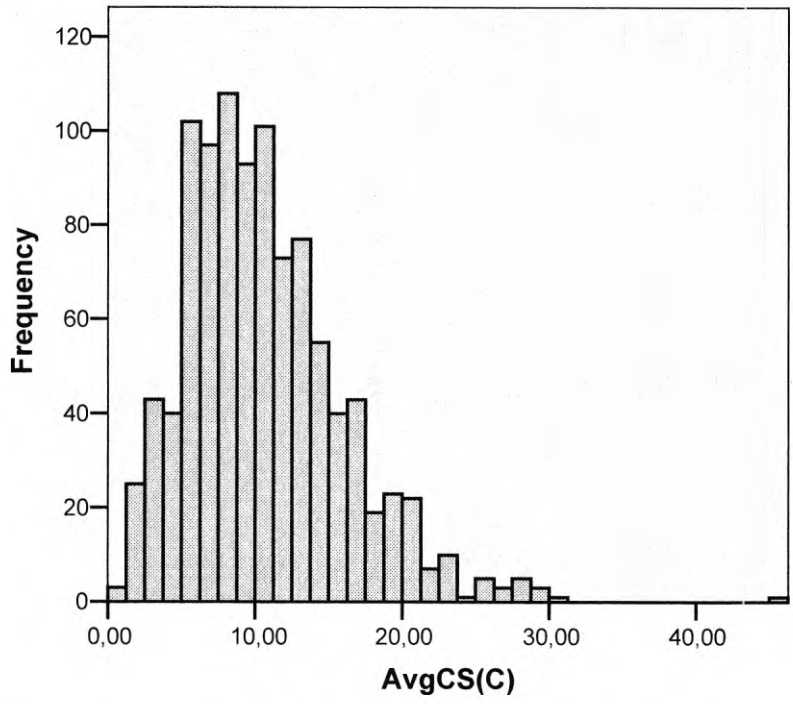


As can be seen from Figure 5-24, the distribution of core documents over clusters is positively skewed. The median cluster size is 4. Approximately 78 percent of all core documents are contained within these clusters and 22 percent in clusters of lesser size.

4.1.2 Coherence and Separation

For the 1000 C1-clusters containing at least three articles, the AvgCS(C) (see equation 4.2) was calculated. The shape of the distribution is summarized in Figure 5-25.

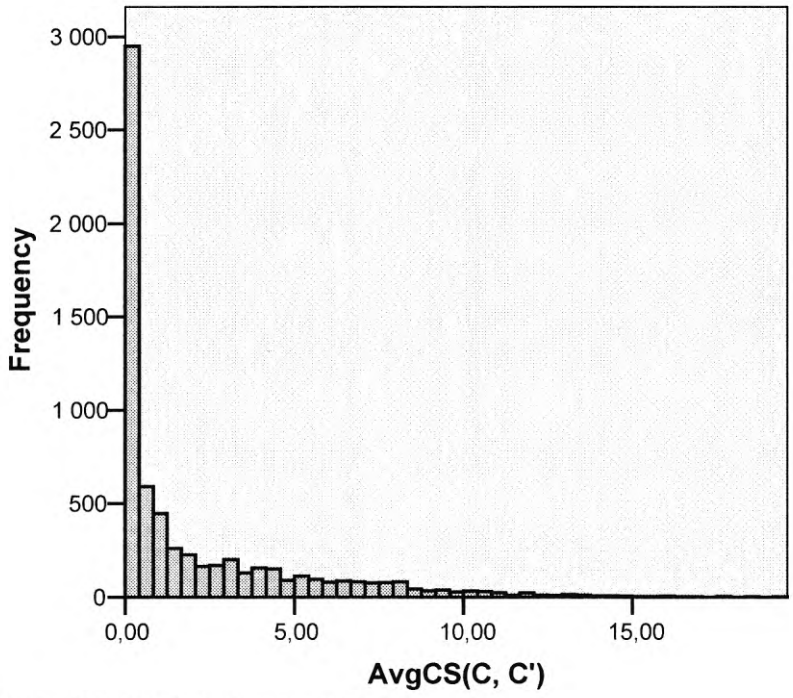
Figure 5-25: The Distribution of Coefficients of AvgCS(C) over 1000 C1 Clusters



As can be seen from Figure 5-25, the distribution is quite symmetrical and the mean AvgCS(C) is 10.59.

With regard to the aspect of isolation, the AvgCS(C, C') (see equation 4.3) between 1000 core document clusters was calculated. Of these, 49 core document clusters were isolated. The shape of the distribution is summarized in Figure 5-26.

Figure 5-26: The Distribution of Coefficients of AvgCS(C, C') over Links between 951 C1-Clusters



Note: 49 isolated clusters are excluded.

As can be seen from Figure 5-26, the distribution is positively skewed. The median of AvgCS(C, C') is 0.65, excluding isolated clusters. Though the median is relatively low, a large number of clusters are strongly connected. Counting links where the distance between C1-clusters > Q3[45] (where the AvgCS(C, C') > 3.11), 1,628 links connecting 706 C1-clusters are found. This shows that a large share of clusters on the C1 level is connected by relatively strong links.

4.1.3 Example of Cluster Fusion on the C1-Level

The fusion of core documents on the C1 level is illustrated by cluster C1/1391. In this cluster, the focus is on SARS[46] and all eleven constituent papers consistently treat this subject. The core documents in this cluster were published in ten different journals and assigned twelve different journal subject categories. The constituent articles are presented with article number, article title, journal title and journal subject category as follows:

i. 27178/ Chest-X-Ray Imaging of Patients with SARs / *Chinese Medical Journal* / **Medicine, General & Internal**

ii. 28525/ Reovirus, Isolated from SARs Patients/ *Chinese Science Bulletin* / **Multidisciplinary Sciences**

[45] Q3 denotes the third quartile, i.e., the score that divides the bottom three quarters of the distribution from the top quarter.

[46] SARS is the acronym for Severe Acute Respiratory Syndrome.

iii. 110241/ Infection-Control for SARs in a Tertiary Neonatal Center/ *Archives of Disease in Childhood*/ **Pediatrics**

iv. 219617/ Description and Clinical Treatment of an Early Outbreak of Severe-Acute-Respiratory-Syndrome (SARs) in Guangzhou, Pr-China/ *Journal of Medical Microbiology*/ **Microbiology**

v. 275806/ A Clinicopathological Study of 3 Cases of Severe Acute Respiratory Syndrome (SARs)/ *Pathology*/ **Pathology**

vi. 333109/ Severe Acute Respiratory Syndrome (SARs) - The Questions Raised by the Management of a Patient in Besancon and Strasbourg/ *Presse Medicale*/ **Medicine, General & Internal**

vii. 383006/ Evaluation of WHO Criteria for Identifying Patients with Severe Acute Respiratory Syndrome Out-of-Hospital - Prospective Observational Study/ *British Medical Journal*/ **Medicine, General & Internal**

viii. 400512/ Severe Acute Respiratory Syndrome-Associated Coronavirus Infection/ *Emerging Infectious Diseases*/ **Immunology; Infectious diseases**

ix. 400574/ Microbiologic Characteristics, Serologic Responses, and Clinical-Manifestations in Severe Acute Respiratory Syndrome, Taiwan/ *Emerging Infectious Diseases*/ **Immunology; Infectious diseases**

x. 490401/ Safe Tracheostomy for Patients with Severe Acute Respiratory Syndrome/ *Laryngoscope*/ **Medicine, Research & Experimental; Otorhinolaryngology**

xi. 527101/ Severe Acute Respiratory Syndrome in Hemodialysis-Patients-A Report of 2 Cases/ *Nephrology Dialysis Transplantation*/ **Transplantation; Urology & Nephrology**

It can be seen from the above list of constituent articles that the problem of SARS is treated with a point of departure in several problem areas like diagnosis (both in vitro as well as in vivo), pathology, clinical treatment and specific clinical problems associated with this syndrome. It would also be interesting to know which cited works connect the core documents in cluster 1391 (see Table 5-11). As the size of this cluster is 11, cited works with a frequency of 11 would be common to all articles in this cluster.

Table 5-11: The Frequency of Works Cited by Articles in Cluster 1391

Frequency	Cited Work
11	Drosten C, 2003, V348, P1967, New Engl J Med
11	Lee N, 2003, V348, P1986, New Engl J Med
9	Peiris JSM, 2003, V361, P1319, Lancet
9	Tsang KW, 2003, V348, P1977, New Engl J Med
8	Ksiazek TG, 2003, V348, P1953, New Engl J Med
7	Poutanen SM, 2003, V348, P1995, New Engl J Med
2	Ho W, 2003, V361, P1313, Lancet
2	Hon KLE, 2003, V361, P1701, Lancet
2	Li TST, 2003, V361, P1386, Lancet
2	Peiris JSM, 2003, V361, P1767, Lancet
2	WHO, 0000, CAS DEF Surv SEV AC
2	WHO, 0000, Cum Numb Rep Prob CA

Note: Only cited works with a frequency > 1 are shown in the table.

Three of the cited works with the highest frequencies are all published in the same journal and in the same issue (*New England Journal of Medicine*, 2003, V348, N20). In addition, they are all about the outbreak of SARS in Hong-Kong in the year 2003:

i. Drosten C et al/ *Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome*

ii. Lee N et al/ *A Major Outbreak of Severe Acute Respiratory Syndrome in Hong-Kong*

iii. Tsang KW et al/ *A Cluster of Cases of Severe Acute Respiratory Syndrome in Hong-Kong*

Studying the compiled reference list of this cluster, the novelty value of the identified research theme on SARS and the identification of a research front issue is clear.

4.2 The Second Fusion Level – C2 Clusters

4.2.1 Clusters and Cluster Sizes

On the basis of the AvgCS(C, C') between C1-clusters containing at least three articles, 6,537 links connecting 951 C1-clusters were applied for an iterated clustering. No threshold of AvgCS(C, C') was applied. The clustering of the 951 C1-clusters resulted in 153 singleton clusters and 212 clusters varying in size between 5 and 97 articles (see Table 5-12). In total, 3,524 core documents were contained in the set of 212 C2-clusters. These were selected for further analysis and fusion.

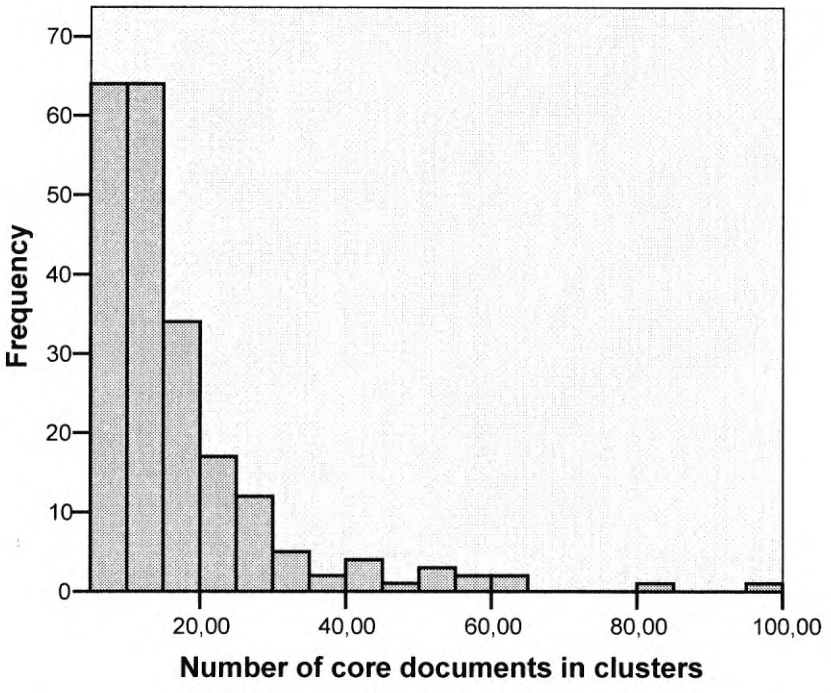
Table 5-12: The Size-Frequency Distribution of C2-Clusters

Size Interval	Frequency
90-99	1
80-89	1
70-79	0
60-69	2
50-59	5
40-49	5
30-39	7
20-29	29
10-19	98
0-9	64
Sum	212

Note: Singleton clusters are excluded and the class interval is 10.

The shape of the distribution is summarized in Figure 5-27.

Figure 5-27: The Distribution of Core Documents over C2-Clusters



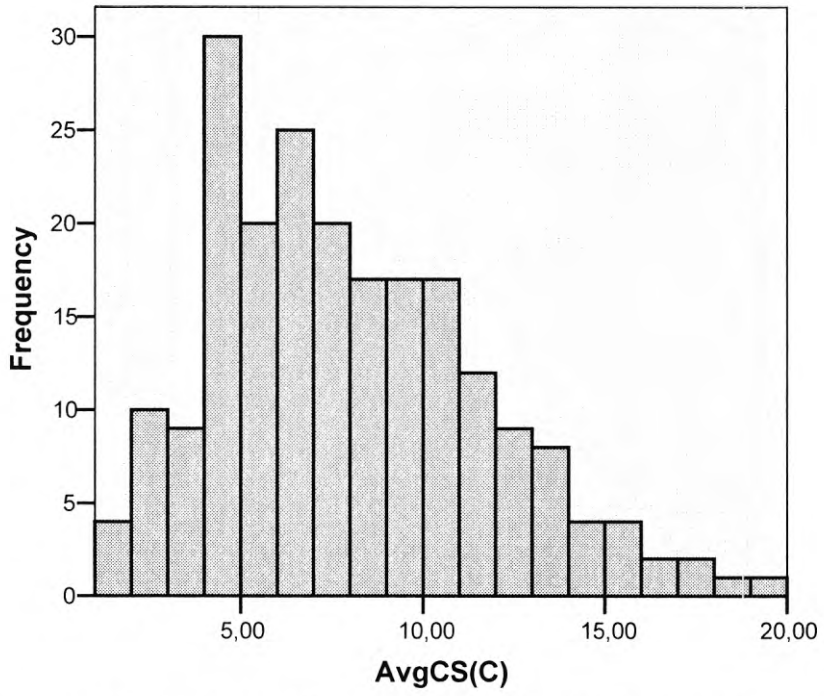
As can be seen from Figure 5-27, the distribution of core documents over 212 clusters is positively skewed. The median cluster size is 13. At the tail of the distribution, nine macro clusters with a size > 50 can be seen. These are foremost large physics clusters, but two are from the bio-medical sciences.[47]

4.2.2 Coherence and Separation

For the 212 C2-clusters, the AvgCS(C) was calculated. The shape of the distribution is summarized in Figure 5-28.

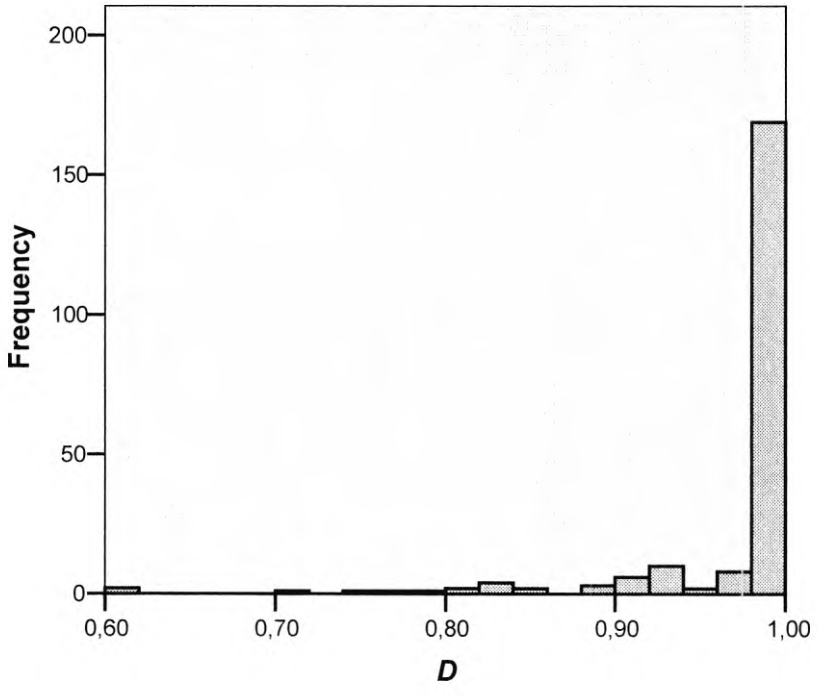
[47] The following fields were presented by macro cluster in accordance to their sizes: Particle physics, Condensed Matter, Crystallography, Applied Physics and Endocrinology & Oncology.

Figure 5-28: The Distribution of Coefficients of AvgCS(C) Over 212 C-2 Clusters



As can be seen from Figure 5-28, the distribution is still rather symmetrical, though the mean AvgCS(C) has dropped from 10.59 on the C1-level to 7.95. *D* (see equation 2.2) was calculated as the density of links between articles in C2-clusters. The shape of the distribution is shown in Figure 5-29.

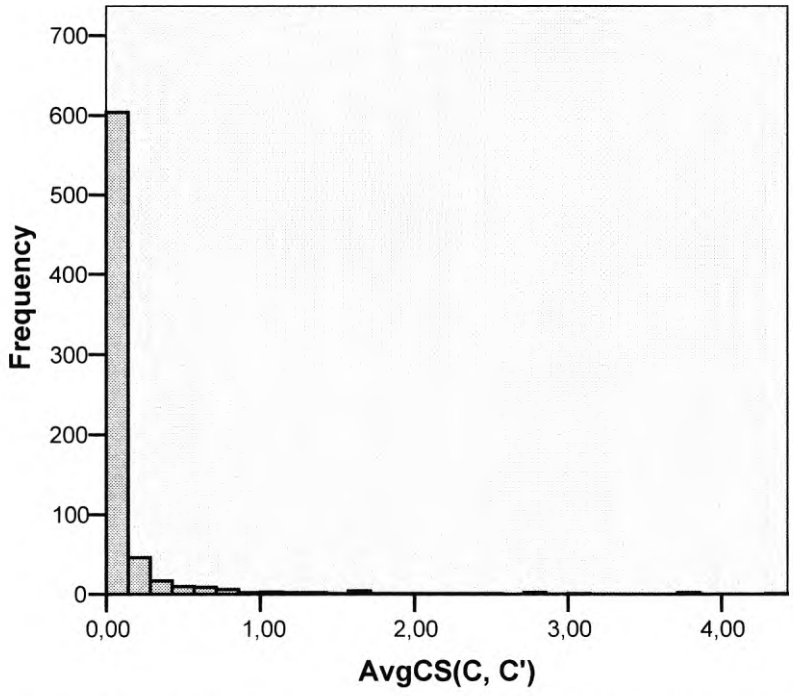
Figure 5-29: The Distribution of Coefficients of *D* over 212 C2-Clusters.



As can be seen from Figure 5-29, the distribution is negatively skewed and most clusters on the C2-level still form complete graphs. The median of *D* was 1.0 (mean = 0.98). The range of *D* was set by a lowest value of *D* of 0.60.

Focusing on the extent to which C2 clusters were separated from one another, the AvgCS(C, C') between C2-clusters were calculated and a total of 23 isolated clusters were found. The shape of the distribution is summarized in Figure 5-30.

Figure 5-30: The Distribution of Coefficients of AvgCS(C, C') over Links between 189 C-2 Clusters



As can be seen from Figure 5-30, the distribution is positively skewed. The median was 0.02. At the tail of the distribution, a few links between C2-clusters indicate that a few relatively strong associations between clusters remain also on this level. Counting links where the distance between C2-clusters > Q3 (where the AvgCS(C, C') > 0.0625), 179 links connecting 145 C2-clusters were found. In all, it means that there is a drastic reduction of strong links and number of connected clusters on this second level of cluster fusion.

4.2.3 Example of Cluster Fusion on the C2-Level

As an example, cluster C1/1391 (see Sub-section 4.1.3 in this section), is fused with four other C1-clusters on the C2-level to C2/170. The fusion is illustrated in Table 5-13.

Table 5-13: The Fusion of Five C1-Clusters to C2/170

AvgCS(C, C')	No. Shared References	C1-Cluster	Cluster Size	C1-Cluster	Cluster Size
5.42	65	1501	4	1513	3
5.14	185	1390	9	1501	4
5.00	135	1390	9	1513	3
4.27	188	1391	11	1501	4
4.03	133	1391	11	1513	3
4.00	392	1390	9	1391	11
3.60	72	1389	5	1501	4
3.56	160	1389	5	1390	9
3.47	52	1389	5	1513	3
3.11	171	1389	5	1391	11

As can be seen from Table 5-13, all clusters are associated with one another with a strength clearly above the median AvgCS(C, C') on the C1-level, forming a complete graph on the C2-level. The mean AvgCS(C) of this graph is 4.25 and *D* is 1.0. Hence, the average connectedness is lower than the mean value for the whole set of C2-clusters though all constituent core documents are bibliographically coupled with one another. Though the average strength of links in the resulting C2 cluster is below the mean for the population of C2-clusters, a clear subject relatedness between constituent core documents and C1-clusters is seen when tabulating and sorting core document titles in accordance to cluster affiliation (see Table 5-14).

Table 5-14: Titles of Articles in Five C1-Clusters Merged to C2/170

1389	Clinical Analysis of 45 Patients with Severe Acute Respiratory Syndrome
1389	The Role of Radiological Imaging in Diagnosis and Treatment of Severe Acute Respiratory Syndrome
1389	Initial Otolaryngological Manifestations of Severe Acute Respiratory Syndrome in Taiwan
1389	A Young Infant with Severe Acute Respiratory Syndrome
1389	Clinical Presentation and Outcome of Severe Acute Respiratory Syndrome in Dialysis Patients
1390	Zcurve-Cov - A New System to Recognize Protein-Coding Genes in Coronavirus Genomes, and Its Applications in Analyzing SARs-Cov Genomes
1390	Prediction of Proteinase Cleavage Sites in Polyproteins of Coronaviruses and Its Applications in Analyzing SARs-Cov Genomes
1390	Maintaining Dental Education and Specialist Dental-Care During an Outbreak of a New Coronavirus Infection - Part 1 - A Deadly Viral Epidemic Begins
1390	Role of China in the Quest to Define and Control Severe- Acute-Respiratory-Syndrome
1390	A Hospital Outbreak of Severe Acute Respiratory Syndrome in Guangzhou, China
1390	An Outbreak of Severe Acute Respiratory Syndrome Among Hospital Workers in a Community-Hospital in Hong-Kong
1390	Epidemiology and Cause of Severe Acute Respiratory Syndrome (SARs) in Guangdong, Peoples-Republic-of-China, in February, 2003
1390	Transmission Dynamics of the Etiologic Agent of SARs in Hong- Kong - Impact of Public-Health Interventions
1390	Children Hospitalized with Severe Acute Respiratory Syndrome- Related Illness in Toronto

Table 5-14 continued…

1391	Severe Acute Respiratory Syndrome-Associated Coronavirus Infection
1391	Microbiologic Characteristics, Serologic Responses, and Clinical-Manifestations in Severe Acute Respiratory Syndrome, Taiwan
1391	Chest-X-Ray Imaging of Patients with SARs
1391	Severe Acute Respiratory Syndrome (SARs) - The Questions Raised by the Management of a Patient in Besancon and Strasbourg
1391	Evaluation of Who Criteria for Identifying Patients with Severe Acute Respiratory Syndrome Out-of-Hospital - Prospective Observational Study
1391	Safe Tracheostomy for Patients with Severe Acute Respiratory Syndrome
1391	Description and Clinical Treatment of an Early Outbreak of Severe-Acute-Respiratory-Syndrome (SARs) in Guangzhou, Pr- China
1391	Reovirus, Isolated from SARs Patients
1391	A Clinicopathological Study of 3 Cases of Severe Acute Respiratory Syndrome (SARs)
1391	Infection-Control for SARs in a Tertiary Neonatal Center
1391	Severe Acute Respiratory Syndrome in Haemodialysis-Patients - A Report of 2 Cases
1501	Clinical-Course and Management of SARs in Health-Care Workers in Toronto - A Case Series
1501	Outcomes and Prognostic-Factors in 267 Patients with Severe Acute Respiratory Syndrome in Hong-Kong
1501	Newly Discovered Coronavirus as the Primary Cause of Severe Acute Respiratory Syndrome
1501	Severe Acute Respiratory Syndrome in a Hemodialysis-Patient
1513	Severe Acute Respiratory-Distress-Syndrome (SARs) - A Critical-Care Perspective
1513	Enteric Involvement of Severe Acute Respiratory Syndrome- Associated Coronavirus Infection
1513	Investigation of a Nosocomial Outbreak of Severe Acute Respiratory Syndrome (SARs) in Toronto, Canada

Note: The first column holds the numbers of constituent C1-clusters in C2/170.

4.3 The Third Fusion Level - C3 Clusters

4.3.1 Clusters and Cluster Sizes

The application of the complete link cluster method on the last level of cluster fusion resulted in a partition where numerous singleton clusters and a few clusters containing two objects were generated only. This means that an upper limit for the application of iterated clustering is found for the proposed method. Still, the question if links between clusters generated on the C2-level are able to form relevant clusters on the last level of cluster fusion needs to be answered. In order to be able to map such links, the between groups average cluster method was applied (see Sub-section 1.4.1 in Chapter 2).

On basis of the computed AvgCS(C, C') between C2 clusters, 189 C2 clusters were partitioned into 92 singleton clusters and 38 clusters containing more than one C2 cluster (see Table 5-15). No threshold of AvgCS(C, C') was applied.[48] The total sum of articles in the 38 clusters was 1,763.

Table 5-15: The Size Frequency Distribution of C2-Clusters

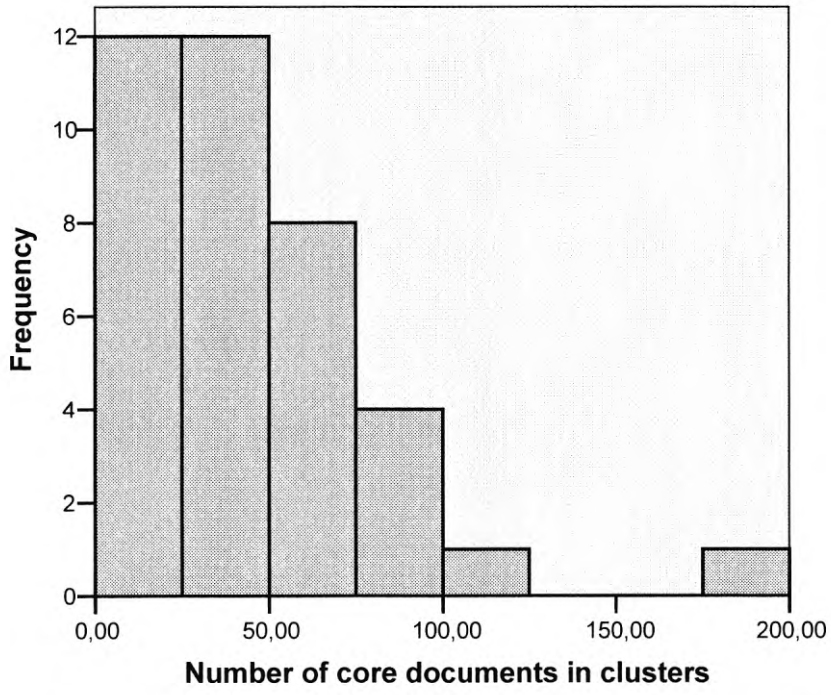
Interval	Frequency
170-199	1
150-169	0
130-149	0
110-129	0
90-109	2
70-89	3
50-69	8
30-49	9
10-29	15
Sum	38

Note: Singleton clusters are excluded and the class interval is 20.

The shape of the distribution is summarized in Figure 5-31.

[48] Extremely low values were regarded as zero associations, hence, the large numbers of singleton clusters on the C3-level (See also Sub-sections 1.3.4 and 1.3.5 in Chapter 5).

Figure 5-31: The Distribution of Core Documents over C3-Clusters

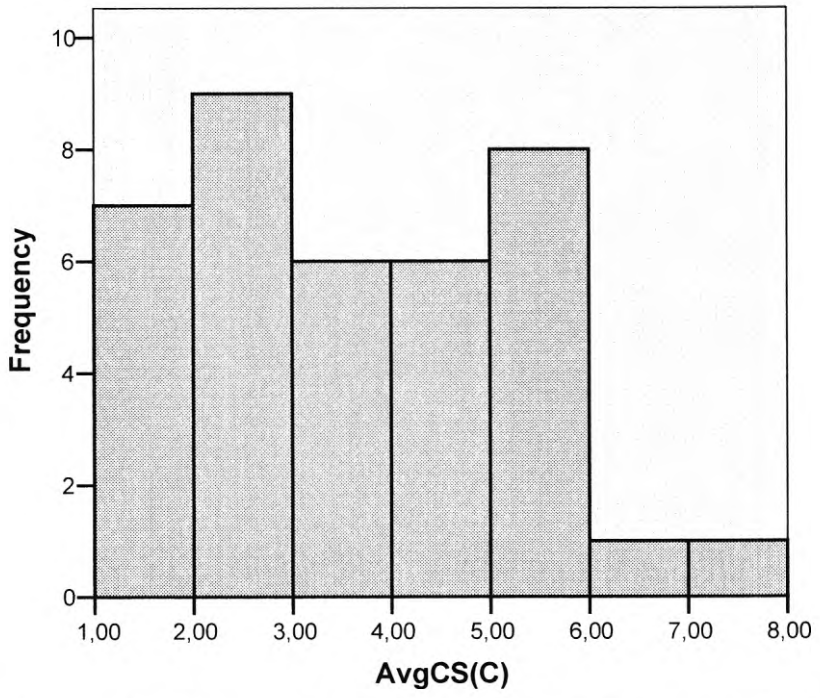


As can be seen from Figure 5-31, the distribution is positively skewed, with most cluster sizes gathered at the lower range of the scale. The median cluster size was 37. Macro clusters ($N \geq 86$) are seen at the higher range of the scale and at the tail. The macro clusters are again from the field of physics (Condensed Matter, Particle Physics, Crystallography and Applied Physics) and from the bio-medical sciences (Oncology-Haematology). One physics cluster (Condensed Matter) builds partly on one of the macro clusters formed at the second fusion level, and one cluster from the bio-medical sciences (Oncology-Haematology) on two of the macro clusters formed at the second fusion level. Otherwise, macro clusters are generated by merging medium and smaller sized C2-clusters.

4.3.2 Coherence and Separation

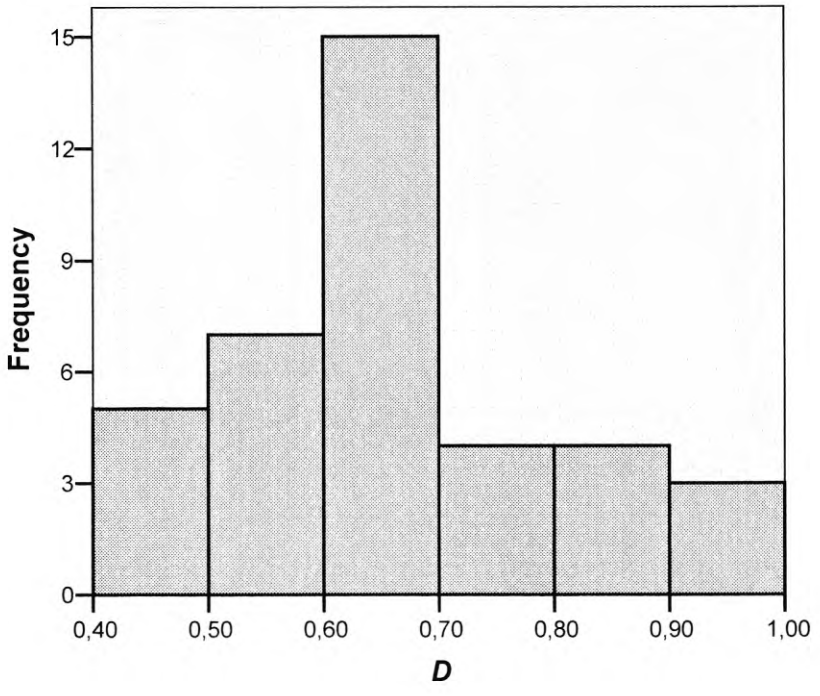
For the 38 C3-clusters, the AvgCS(C) was calculated. The shape of the distribution is summarized in Figure 5-32.

Figure 5-32: The Distribution of Coefficients of AvgCS(C) over 38 C3-Clusters



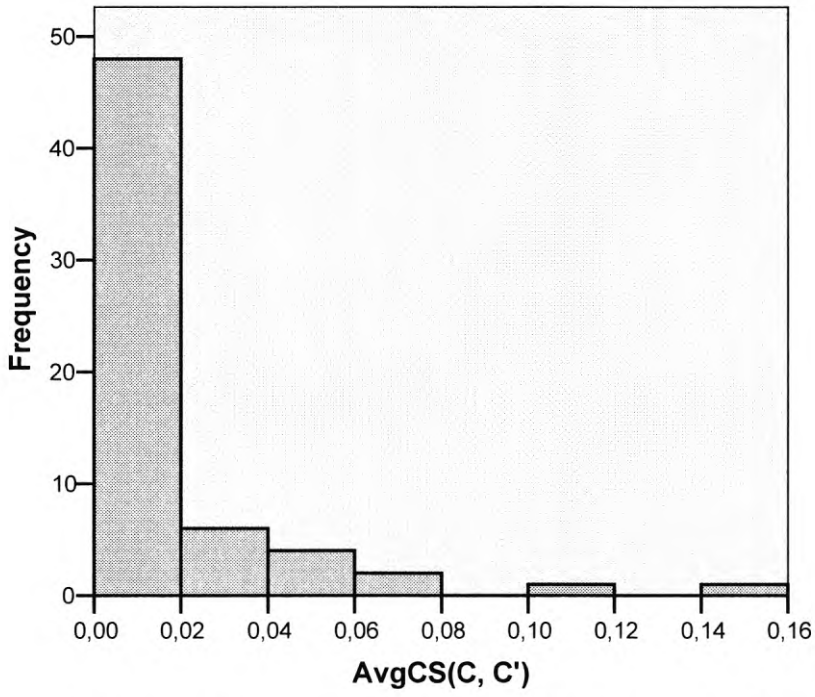
As can be seen from Figure 5-32, the distribution is almost rectangular. Moving up to this level of cluster fusion, the mean AvgCS(C) drops from 7.95 to 3.64. Likewise, the density of links between core documents *D* makes a drop from near the maximum value to a mean of 0.66 (md = 0.64) (see Figure 5-33).

Figure 5-33 : The Distribution of Coefficients of *D* over 38 C3-Clusters



With regard to the aspect of isolation, the AvgCS(C, C') between 38 C3-clusters was calculated. The shape of the distribution is summarized in Figure 5-34.

Figure 5-34: The Distribution of Coefficients of AvgCS(C, C') over Links between 38 C3-Clusters



As can be seen from Figure 5-34, the distribution is positively skewed. Six C3-clusters are isolated. The median AvgCS(C, C') between the remaining 32 C3-clusters has dropped from 0.02 to 0.002 and mostly spurious links between C3-clusters remain.

4.3.3 Example of Cluster Fusion on the C3-Level

On the final level of cluster fusion, cluster C2/170 formed on the second level of cluster fusion is merged with two other C2-clusters to C3/3, with a total number of 61 core documents (see Table 5-16).

Table 5-16: The Fusion of Three C2-Clusters to C3/3

AvgCS(C, C')	No. Shared References	C2-Cluster	Cluster Size	C2-Cluster	Cluster Size
0.25	129	87	16	170	32
0.10	20	87	16	171	13
2.48	1030	170	32	171	13

As can be concluded from Table 5-16 above, cluster C2/87 presents the most distant (or dissimilar) vertex in a three-edged graph. Hence, from some aspect of similarity, it should deviate from the two other C2-clusters. This is,

however, not clearly reflected by the mix of journal subject categories assigned to the following core documents in the C2-clusters:

C2/ 87: infectious diseases, clinical microbiology;

C2/ 170: general & internal medicine, infectious diseases, biochemistry ; pediatrics; urology & nephrology; and

C2/ 171: biochemistry & molecular biology, clinical chemistry, microbiology, virology.

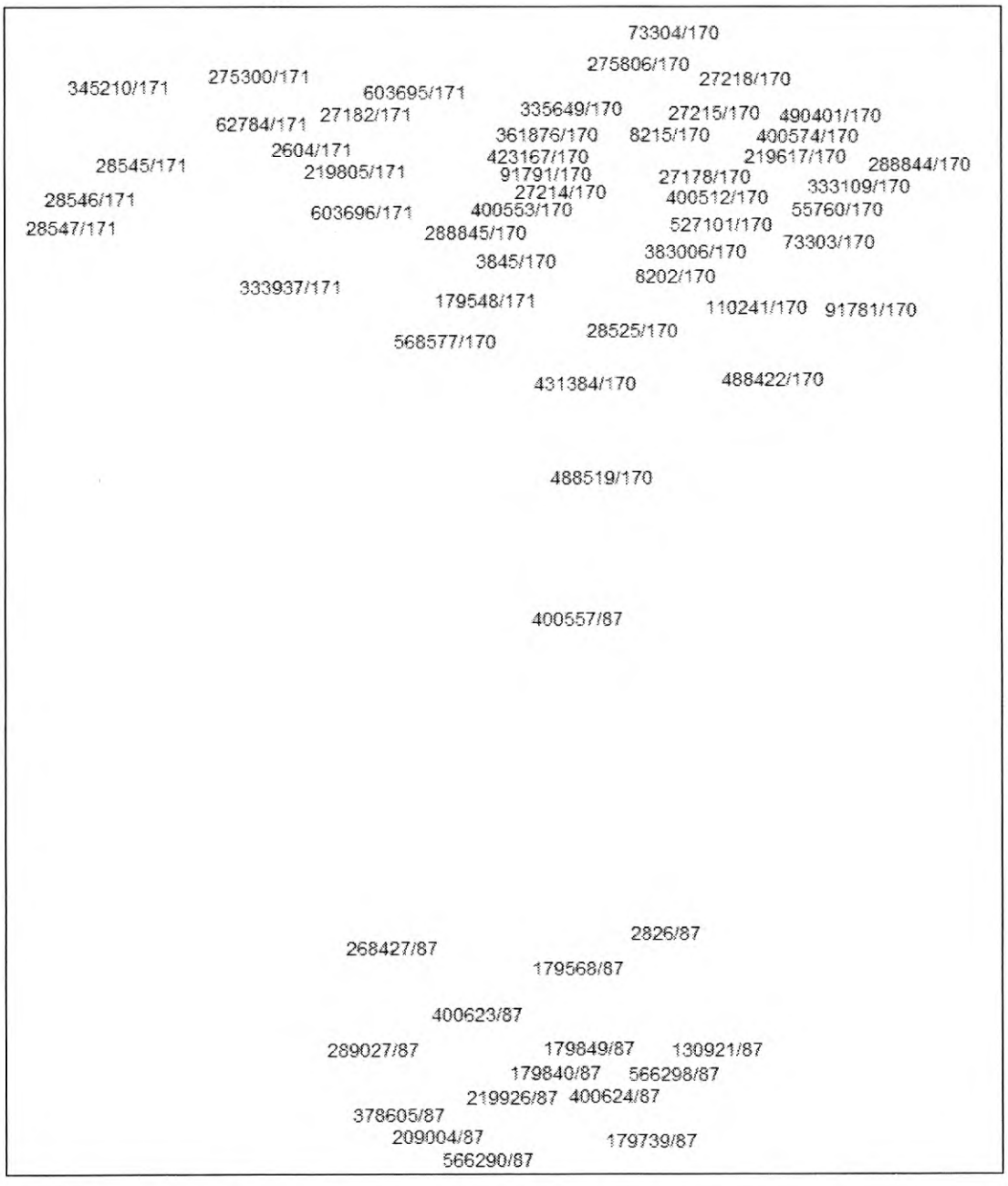
Approximately 11 different disciplines are more or less associated with the research theme(s) of C3/3.[49] All three C2 clusters constituting C3/3 were complete graphs, with regard to links between core documents. The values of Avg(C) ranged between 3.54 and 5.32. The values of *D* and AvgCS(C) for the graph of C3/3 were 0.62 and 2.30 respectively. Hence, the coherence was lower than for the average C3 cluster.

At this level of cluster fusion, the structure is more complex and worthy of a more thorough analysis. As such, a "top-bottom" interpretation is suggested.

Beginning this analysis from the "top", all links between core documents in C3/3 are displayed in a two dimensional plane by MDS with an acceptable value of stress (see Figure 5-35). In this graph, the constituent C2 clusters are clearly discernable. Clusters C2/171 and C2/170 are configured in the upper part of the map and cluster C2/87 in the lower part with the core document 400557 in an intermediate position.

[49] The exact number of disciplines will not be given by assigned journal subject categories as these are journal classifications, covering the scope of journals, not the scope of the individual paper.

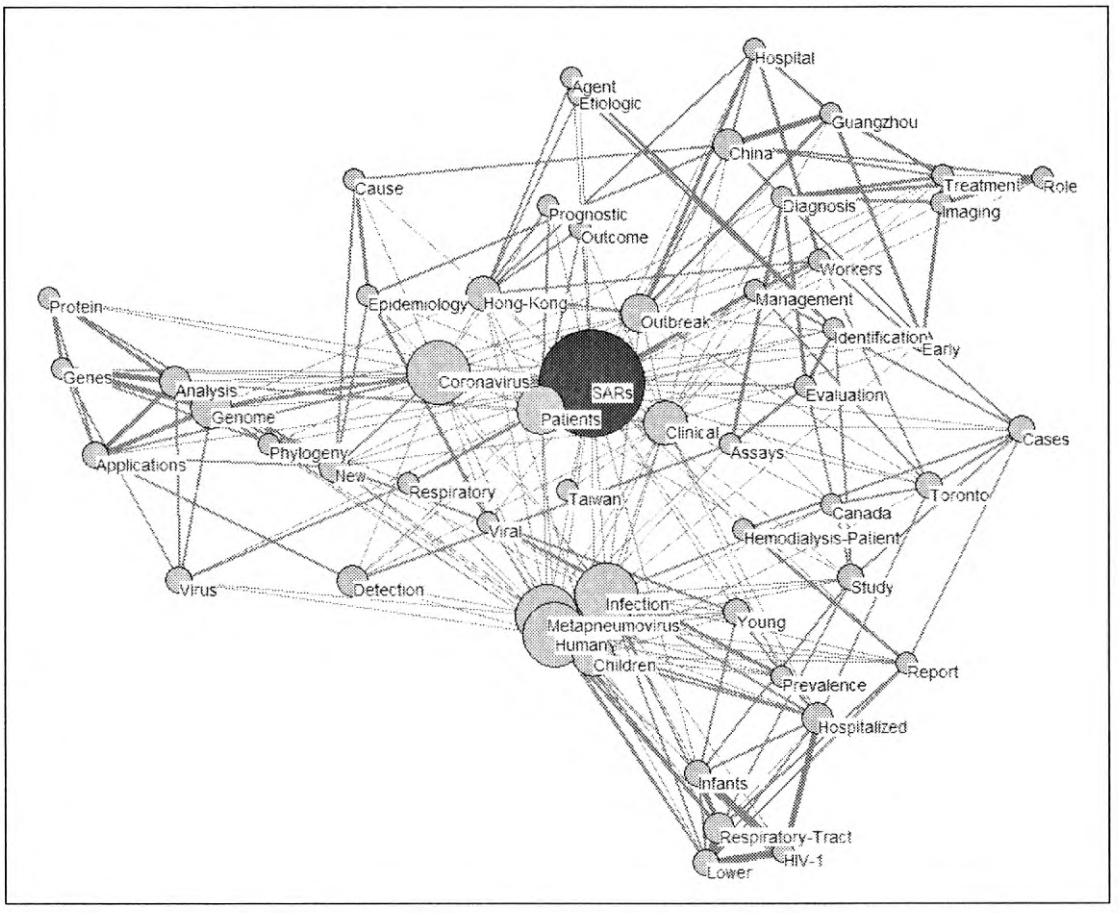
Figure 5-35: The Configuration of C3/3 Constituted by C2/87, C2/ 170 & C2/ 171



Note: i. Numbers on the map correspond to document numbers and C2-cluster numbers.
ii. Kruskal's stress is 0.11

Applying MDS to display the associations between significant terms (title words) according to their co-occurrence in titles, an overview sketch of the topic content in cluster C3/3 is arrived at (see Figure 5-36).

Figure 5-36: MDS Display of the Co-occurrence of Title Words in the Core Documents of C3/3



Note: i. The lowest frequency of term occurrence allowed for is 2.

ii. The width of links is corresponding to the strength of similarity between terms as measured by the Jaccard index (see Equation 2.4).

iii. Circle sizes correspond to the frequency of occurrence.

iv. Kruskal's stress is 0.10.

As can be seen from Figure 5-36, the focus of core documents in cluster C3/3 is on infectious diseases caused by viruses (in particular SARS). To begin with, the configuration in terms of term size-position could be examined. The term "SARs" has a central position and, as indicated by the circle size, it is the most frequent term. Radiating out from "SARs", different dimensions associated with diseases caused by viruses can be discerned:

i. a time-geography dimension (outbreak; early; epidemiology; Hong-Kong; China; Guangzhou; Taiwan; Canada; Toronto);

ii. a clinical dimension (metapneumovirus; human; infection; children; young; infants; lower; respiratory tract; HIV 1; hospitalized; prevalence); and

iii. a dimension of the genetics of viruses (analysis; genome; phylogeny; new; protein; genes; application; virus).

Zooming in on particular C2-clusters, different aspects of cluster C3/3 could be reflected. Beginning with cluster C2/87 (located on the lower part of the map in Figure 5-35), this cluster is constituted by two C1 clusters and the total number of core documents was 16. Studying the titles of core documents constituting C2/87, the subject homogeneity is obvious as they are all on *Human Metapneumovirus*. Also, a preliminary explanation of the intermediate role of core document 400557 (see Figure 5-35) is that it associates this virus with SARS. SARS is for all associated with a corona virus (SARS-CoV), but also with the human metapneumo virus, though to a lesser degree (see Table 5-17).

Table 5-17: Core Document Titles in C2/87

2826 / Human Metapneumovirus-Associated Lower Respiratory-Tract Infections Among Hospitalized Human-Immunodeficiency-Virus Type-1 (HIV-1)-Infected and HIV-1-Uninfected African Infants
130921 / Prevalence and Clinical Symptoms of Human Metapneumovirus Infection in Hospitalized-Patients
179568 / Human Metapneumovirus Infection in the Canadian Population
179739 / Comparative-Evaluation of Real-Time PCR Assays for Detection of the Human Metapneumovirus
179840 / Human Metapneumovirus Associated with Respiratory-Tract Infections in a 3-Year Study of Nasal Swabs from Infants in Italy
179849 / High Prevalence of Human Metapneumovirus Infection in Young- Children and Genetic-Heterogeneity of the Viral Isolates
209004 / Human Metapneumovirus Infections in Young and Elderly Adults
219926 / Seroprevalence of Human Metapneumovirus in Japan
268427 / Effects of Human Metapneumovirus and Respiratory Syncytial Virus-Antigen Insertion in 2 3'-Proximal Genome Positions of Bovine /Human Parainfluenza Virus Type-3 on Virus-Replication and Immunogenicity
289027 / Human Metapneumovirus Infection in the United-States - Clinical-Manifestations Associated with a Newly Emerging Respiratory-Infection in Children
378605 / Human Metapneumovirus in a Hematopoietic Stem-Cell Transplant Recipient with Fatal Lower Respiratory-Tract Disease
400557 / Human Metapneumovirus Detection in Patients with Severe Acute Respiratory Syndrome
400623 / Children with Respiratory-Disease Associated with Metapneumovirus in Hong-Kong
400624 / Human Metapneumovirus Infections in Hospitalized Children 219926 / Seroprevalence of Human Metapneumovirus in Japan
566290 / Human Metapneumovirus Infection in Thai Children

Clinical findings and studies of prevalence with regard to human metapneumo virus (with some emphasis on children) are presented by the core document titles. Looking at titles and journal subject category assignments of journals in which these core documents were published, the disciplinary structure leans towards general (internal) medicine (infectious diseases) but there is also a contribution from basic medical sciences (immunology, microbiology) (see Table 5-18).

Table 5-18: Journal Titles and Assigned Journal Subject Categories Corresponding to Core Documents in C2/87 In C3/3

Journal Title	Journal Subject Categories
(1) Clinical Infectious Diseases	Immunology; infectious diseases; microbiology
(3) Emerging Infectious Diseases	Immunology; infectious diseases
(2) Journal of Infectious Diseases	Infectious diseases
(2) Scandinavian Journal of Infectious Diseases	Infectious diseases
(4) Journal of Clinical Microbiology	Microbiology
(1) Journal of Medical Virology	Virology
(1) Journal of Virology	Virology
(1) Pediatrics	Paediatrics
(1) Bone Marrow Transplantation	Oncology; hematology; immunology; transplantation

Note: Numbers in brackets correspond to the frequency of articles published in a journal.

The next constituent C2-cluster to be studied is C2/171 (located on upper left quadrant in Figure 5-35) which is formed by three C1 clusters and 13 core documents. Table 5-19 gives the core document titles in this cluster.

Table 5-19: Core Document Titles in C2/171

2604 / Quantitative-Analysis and Prognostic Implication of SARs Coronavirus RNA in the Plasma and Serum of Patients with Severe Acute Respiratory Syndrome
179548 / Evaluation of Reverse Transcription-PCR Assays for Rapid Diagnosis of Severe Acute Respiratory Syndrome-Associated with a Novel Coronavirus
219805 / Early Events of SARs Coronavirus Infection in Vero Cells
27182 / Establishment of a Fluorescent Polymerase-Chain-Reaction Method for the Detection of the SARs-Associated Coronavirus and Its Clinical-Application
28545 / Design and Application of 60Mer Oligonucleotide Microarray in SARs Coronavirus Detection
28546 / Molecular Phylogeny of Coronaviruses Including Human SARs-Cov
28547 / Phylogeny of SARs-Cov as Inferred from Complete Genome Comparison
333937 / Activation of Ap-1 Signal-Transduction Pathway by SARs Coronavirus Nucleocapsid Protein
345210 / Genomic Characterization of the Severe-Acute-Respiratory- Syndrome Coronavirus of Amoy Gardens Outbreak in Hong-Kong
603695 / Coronavirus in Severe Acute Respiratory Syndrome (SARs)
603696 / Severe Acute Respiratory Syndrome - Identification of the Etiologic Agent
62784 / Mutation Analysis of 20 SARs Virus Genome Sequences - Evidence for Negative Selection in Replicase Orf1B and Spike Gene
275300 / The Crystal-Structures of Severe Acute Respiratory Syndrome Virus Main Protease and Its Complex with an Inhibitor

As can be seen from Table 5-19, this cluster focuses exclusively on corona virus and SARS. The emphasis is on the analysis of the genetic

characterization and on methods for the detection and description of the viruses (e.g. laboratory methods, isolation and cultivation). The clinical focus seen in cluster C2/87 is thus replaced with more basic research in the virus causing SARS. This is also reflected by the composition of the set of publishing journals and journal subject categories of this cluster, where the contribution from chemistry and biochemistry is salient (see Table 5-20).

Table 5-20: Journal Titles and Assigned Journal Subject Categories Corresponding to Core Documents in Cluster C2/171

Journal Title	Journal Subject Categories
(3) Chinese Science Bulletin	Multidisciplinary sciences
(2) Trends in Molecular Medicine	Biochemistry & molecular biology; cell biology; medicine, research & experimental
(1) ACTA Pharmacologica Sinica	Chemistry, multidisciplinary; pharmacology & pharmacy
(1) Biochemical and Biophysical Research Communications	Biochemistry & molecular biology; biophysics
(1) Chinese Medical Journal	Medicine, general & internal
(1) Clinical Chemistry	Medical laboratory technology
(1) Journal of Clinical Microbiology	Microbiology
(1) Journal of Medical Virology	Virology
(1) Lancet	Medicine, general & internal
(1) Proceedings of the National Academy of Sciences of the United States of America	Multidisciplinary sciences

Note: Numbers in brackets correspond to the frequency of papers published in a journal.

Lastly, the largest C2 cluster (i.e. C2/170) constituting C3/3 was studied (located on upper right quadrant in Figure 5-35). This cluster was formed by three C1 clusters and 32 core documents. In this cluster, several case studies as well as clinical aspects on diagnosis and prevention are reported, and the overall focus is again on clinical aspects of SARS (see Table 5-21).

Table 5-21: Core Document Titles in Cluster C2/170

3845 / Severe Acute Respiratory-Distress-Syndrome (SARs) - A Critical-Care Perspective
8202 / Investigation of a Nosocomial Outbreak of Severe Acute Respiratory Syndrome (SARs) in Toronto, Canada
8215 / Clinical-Course and Management of SARs in Health-Care Workers in Toronto - A Case Series
27178 / Chest-X-Ray Imaging of Patients with SARs
27214 / A Hospital Outbreak of Severe Acute Respiratory Syndrome in Guangzhou, China
27215 / Clinical Analysis of 45 Patients with Severe Acute Respiratory Syndrome
27218 / The Role of Radiological Imaging in Diagnosis and Treatment of Severe Acute Respiratory Syndrome
28525 / Reovirus, Isolated from SARs Patients
55760 / Initial Otolaryngological Manifestations of Severe Acute Respiratory Syndrome in Taiwan
73303 / Severe Acute Respiratory Syndrome in a Hemodialysis-Patient
73304 / Clinical Presentation and Outcome of Severe Acute Respiratory Syndrome in Dialysis Patients
91781 / Outcomes and Prognostic-Factors in 267 Patients with Severe Acute Respiratory Syndrome in Hong-Kong
91791 / An Outbreak of Severe Acute Respiratory Syndrome Among Hospital Workers in a Community-Hospital in Hong-Kong
110241 / Infection-Control for SARs in a Tertiary Neonatal Center
219617 / Description and Clinical Treatment of an Early Outbreak of Severe-Acute-Respiratory-Syndrome (SARs) in Guangzhou, Pr- China
275806 / A Clinicopathological Study of 3 Cases of Severe Acute Respiratory Syndrome (SARs)
288844 / A Young Infant with Severe Acute Respiratory Syndrome
288845 / Children Hospitalized with Severe Acute Respiratory Syndrome- Related Illness in Toronto
333109 / Severe Acute Respiratory Syndrome (SARs) - The Questions Raised by the Management of a Patient in Besancon and Strasbourg
335649 / Maintaining Dental Education and Specialist Dental-Care During an Outbreak of a New Coronavirus Infection - Part 1 - A Deadly Viral Epidemic Begins
361876 / Zcurve-Cov - A New System to Recognize Protein-Coding Genes in Coronavirus Genomes, and Its Applications in Analyzing SARs-Cov Genomes
383006 / Evaluation of WHO Criteria for Identifying Patients with Severe Acute Respiratory Syndrome Out-of-Hospital - Prospective Observational Study
400512 / Severe Acute Respiratory Syndrome-Associated Coronavirus Infection
400553 / Role of China in the Quest to Define and Control Severe- Acute-Respiratory-Syndrome
400574 / Microbiologic Characteristics, Serologic Responses, and Clinical-Manifestations in Severe Acute Respiratory Syndrome, Taiwan
423167 / Prediction of Proteinase Cleavage Sites in Polyproteins of Coronaviruses and Its Applications in Analyzing SARs-Cov Genomes
431384 / Enteric Involvement of Severe Acute Respiratory Syndrome- Associated Coronavirus Infection
488422 / Epidemiology and Cause of Severe Acute Respiratory Syndrome (SARs) in Guangdong, Peoples-Republic-of-China, in February, 2003
488519 / Newly Discovered Coronavirus as the Primary Cause of Severe Acute Respiratory Syndrome
490401 / Safe Tracheostomy for Patients with Severe Acute Respiratory Syndrome
527101 / Severe Acute Respiratory Syndrome in Hemodialysis-Patients - A Report of 2 Cases
568577 / Transmission Dynamics of the Etiologic Agent of SARs in Hong- Kong - Impact of Public-Health Interventions

Looking at the distribution of journal subject categories and journal titles, several medical disciplines and sub-disciplines are represented with an emphasis on general medicine (see Table 5-22).

Table 5-22: Journal Titles and Assigned Journal Subject Categories Corresponding to Core Documents in Cluster C2/170

Journal Title	Journal Subject Categories
(5) Chinese Medical Journal	Medicine, general & internal
(3) Emerging Infectious Diseases	Immunology; infectious diseases
(2) American Journal of Kidney Diseases	Urology & nephrology
(2) Annals of Internal Medicine	Medicine, general & internal
(2) Canadian Medical Association Journal	Medicine, general & internal
(2) Lancet	Medicine, general & internal
(2) Paediatrics	Paediatrics
(1) Archives of Disease in Childhood	Paediatrics
(1) Archives of Otolaryngology-Head & Neck Surgery	Otorhinolaryngology; surgery
(1) British Dental Journal	Dentistry, oral surgery & medicine
(1) British Medical Journal	Medicine, general & internal
(1) Chinese Science Bulletin	Multidisciplinary sciences
(1) Critical Care Medicine	Critical care medicine
(1) FEBS Letters	Biochemistry & molecular biology; biophysics; cell biology
(1) Gastroenterology	Gastroenterology & hepatology
(1) Journal of Medical Microbiology	Microbiology
(1) Laryngoscope	Medicine, research & experimental; otorhinolaryngology
(1) Nephrology Dialysis Transplantation	Transplantation; urology & nephrology
(1) Pathology	Pathology
(1) Presse Medicale	Medicine, general & internal
(1) Science	Multidisciplinary sciences

Note: Numbers in brackets correspond to the frequency of papers published in a journal.

It is to be noted that the different MDS maps (i.e. Figure 5-35 and Figure 5-36) match surprisingly well, though the first map groups papers according to shared references and the second map title words according to their co-occurrence frequency in titles in core documents. Hence, C2/87 seems to correspond to the lower part of the "title word" map, C2/171 to the left-middle-upper part and C2/170 to the right-middle-upper part.

It seems clear that the C2-clusters are subject consistent and a common denominator (SARS) can be identified. The interdisciplinary character on all levels (C1 to C3) is obvious. The merging of the three C2-clusters on the C3 level thus connects research in two different viruses and the pathology of

diseases caused by them, genesis of agents and corresponding clinical research and observations. The merging of C2/87 with the other two clusters could, however, be questioned. Though a least common denominator (SARS) exists, both the measured distances and the assessed cognitive distances between C2-clusters showed that there is a need for the separation of cluster C2/87 from the other C2-clusters. Nevertheless, the associations of the clusters could be of interest as they all gather around a common problem, though from different perspectives.

4.4 Field Experts' Evaluations of Four Cases of Iterated Clustering

Four cases of iterated clustering were presented to four different field experts for evaluation. The design of this experiment aimed at finding clusters on the last level of cluster fusion from the three major science fields: physics, chemistry and bio-medicine for the evaluation. C3-clusters from these fields were then matched against profiles of researchers and when a match occurred, a preliminary choice of cluster was made. If a researcher was available to do the evaluation, a final choice of cluster was made. In order to approximate the greater impact of physics on the composition of the database underlying the experiments in Case 4, two cases from the field of physics, and one each from the other two fields were selected. In all, a total of 154 core documents were evaluated.

Corresponding to each case is a C3-cluster, which is assumed to reflect research themes with a cognitive linkage to one another. The field experts were asked to assess the relevance of cluster composition on all three levels of cluster fusion, i.e. C1 to C3 (see Sub-section 2.1 in Chapter 4). In order to illustrate the field experts' evaluations, the internal structure of each C3-cluster is visualized by mapping links between constituent core documents. It is to be noted that there exists a scale difference between maps and they are not directly comparable with each other.

The results of the evaluations are given below.

4.4.1 Cluster C3/12, "Human Genetics and Disease"

This C3-cluster contained 53 core documents distributed over three C2-clusters and 11 C1-clusters as follows:

C2/45 : C1/616; C1/1003
C2/46 : C1/1171; C1/1297; C1/1170; C1/1172
C2/210: C1/9; C1/1163; C1/1168; C1/1184; C1/1286

The focus in this cluster is generally on human genetics and disease. A total of 21 different journal subject categories were assigned the journals in which core documents in this cluster were published (see Table 5-23).[50]

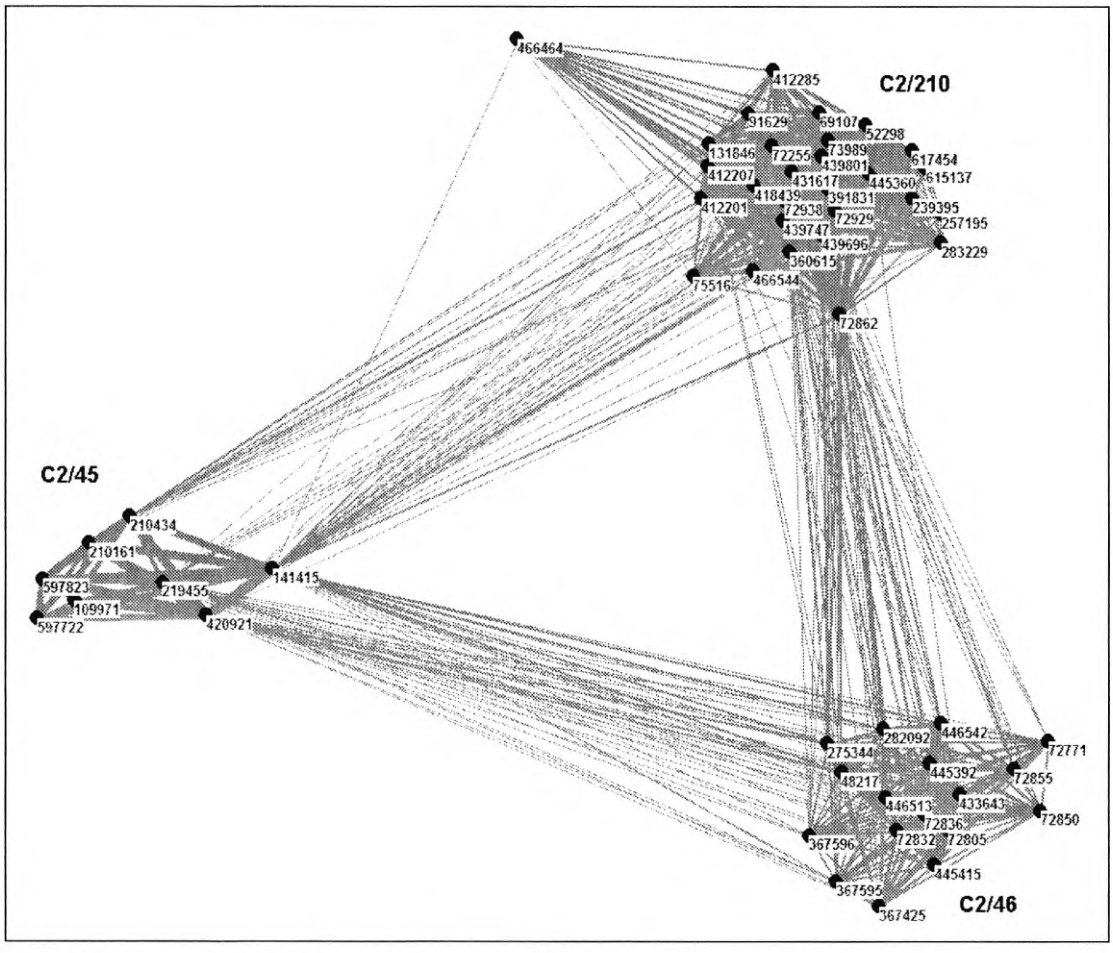
[50] The exact number of contributing disciplines will not be given by assigned journal subject categories as these are journal classifications, covering the scope of journals, not the scope of the individual article.

Table 5-23: The Frequency Distribution of Journal Subject Categories in Cluster C3/12

Frequency	Journal Subject Categories
22	Genetics & Heredity
13	Gastroenterology & Hepatology
5	Dermatology
4	Immunology
3	Biochemical Research Methods
3	Biochemistry & Molecular Biology
3	Biotechnology & Applied Microbiology
3	Cell Biology
3	Computer Science, interdisciplinary applications
3	Mathematics, interdisciplinary applications
3	Multidisciplinary Sciences
3	Statistics & Probability
2	Pathology
1	Nutrition & Dietetics
1	Ophthalmology
1	Pediatrics
1	Pharmacology & Pharmacy
1	Public, Environmental & Occupational Health
1	Respiratory System
1	Rheumatology

In the graph representing C3/12, each of the three C2-clusters are clearly depicted and demarcated as can be seen from Figure 5-37. The density D of the graph was 0.52 and the AvgCS(C) 2.85. Hence, both values of cluster coherence were below the average.

Figure 5-37: The Configuration of Core Documents in Cluster C3/12



Note: Kruskal's stress is 0.06.

The field expert's opinion was that core documents in C1 and C2 clusters were consistently subject related, with two exceptions. The first exception is article 466464 in C2/210/C1/1184 which seemed to have a too general topic content in relation to the pronounced focus in C2/210 on inflammatory bowel-diseases. This was in agreement with its more peripheral position on the map. The second exception was core document 283229 in C2/210/C1/1163, which the expert assumed to be relevant, but with some uncertainty as the title was not exhaustive enough. The field expert renounced judging the relevance of merging disease-gene-mapping methods (C2/46), with research in genetic aspects of psoriasis (C2/45) and inflammatory bowel-disease (C2/210).

4.4.2 Cluster C3/19: "Chemistry"

This cluster contained 25 core documents distributed over two C2-clusters and five C1-clusters as follows:

C2/111: C1/1189; C1/1263
C2/191: C1/81; C1/406; C1/1394

All core documents but one in C3/19 pertained to the field of chemistry and the composition of contributing disciplines varied, though with an emphasis on organic chemistry. A total of five different journal subject categories were assigned the journals in which core documents in this cluster were published (see Table 5-24).[51]

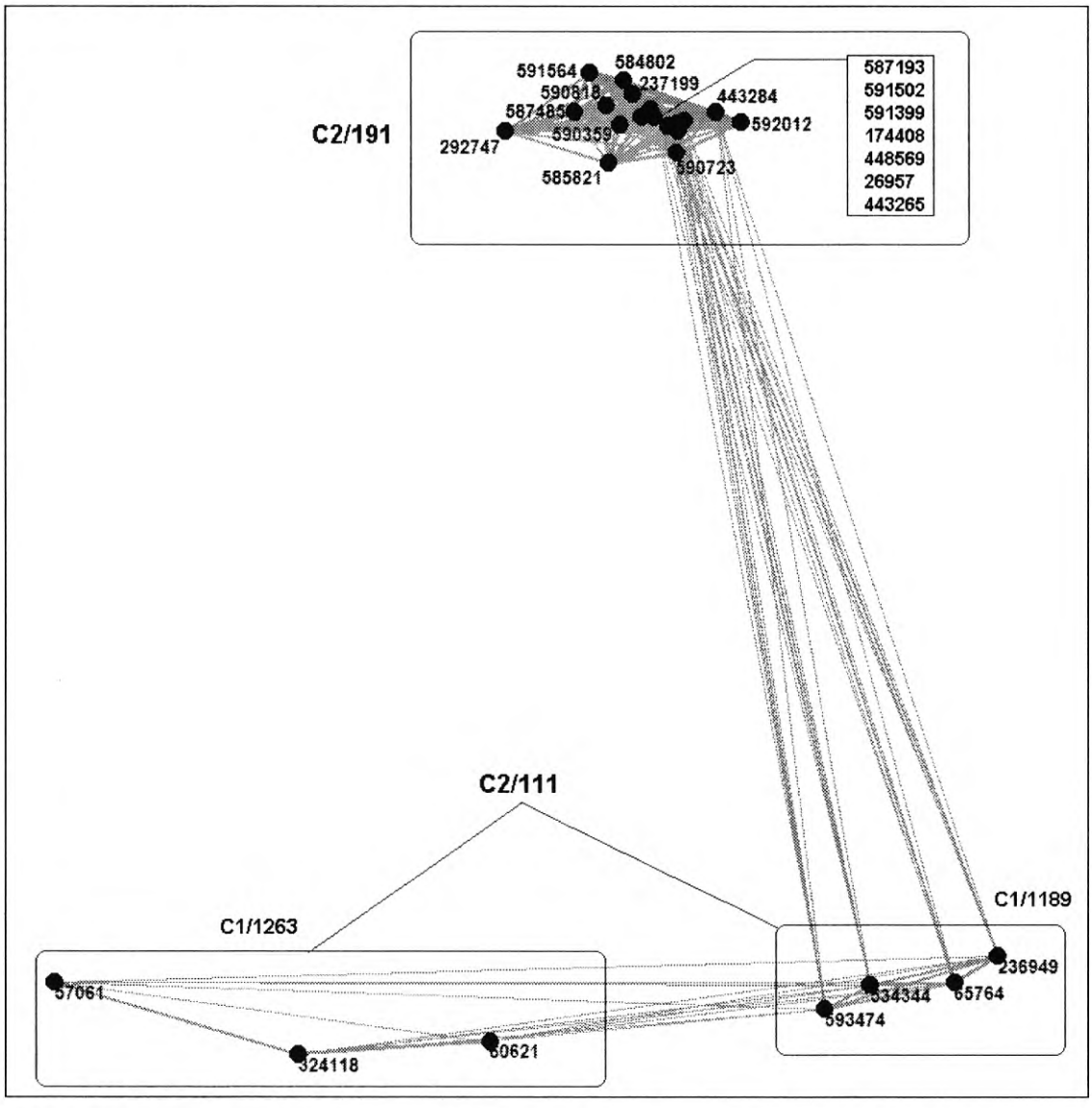
Table 5-24: The Frequency Distribution of Journal Subject Categories in Cluster C3/27

Frequency	Journal Subject Categories
17	Chemistry, Organic
6	Chemistry, Multidisciplinary
1	Chemistry, Inorganic & Nuclear
1	Cell Biology
1	Chemistry, Applied

In the graph depicting C3/19, the composition of C2/191 is visualized as a compact cluster whereas C2/111 is a looser construct, and there exists no links between C1/1263 and C2/111 (see Figure 5-38). In spite of the latter, the coherence of C3/19 is above the average with 5.32 for the AvgCS(C) and 0.67 for *D*.

[51] The exact number of contributing disciplines will not be given by assigned journal subject categories as these are journal classifications, covering the scope of journals, not the scope of the individual article.

Figure 5-38: The Configuration of Core Documents in Cluster C3/19



Note: i. Due to the compactness of C2/191, seven labels representing articles could not be fitted to mark corresponding circles of cluster C2/191 and are presented in the nearby table in the map.
ii. Kruskal's stress was 0.02.

According to the field expert, no misplaced core documents were found on the C1-level. However, C2/111/C1/1263 was found to be more diverse than cluster C2/111/C1/1189, which is reflected by the configuration in the map where C1/1263 form a looser structure. On the C2 level, C2/111 was considered to be subject consistent in terms of a research focus common to the constituent C1 clusters. As for C2/191, the partition in C1-clusters appeared artificial to the field expert and C2/191 was better regarded as one cluster, which is reflected by this cluster's compactness, as seen in the map. Regarding the merging of the C2 clusters, no clear subject relationship between them was obvious.

4.4.3 Cluster C3/27: "Bose-Einstein Condensation"

This cluster contained 54 core documents distributed over four C2-clusters and 13 C1-clusters as follows:

C2/140: C1/367; C1/555
C2/141: C1/459; C1/557; C1/578
C2/143: C1/353; C1/362; C1/439; C1/552
C2/144: C1/352; C1/359; C1/454; C1/643

Articles in this cluster pertain to research areas of optical, atomic & molecular physics and the major focus is on Bose-Einstein condensation.[52] A total of five different journal subject categories were assigned the journals in which core documents in this cluster were published (see Table 5-25).[53]

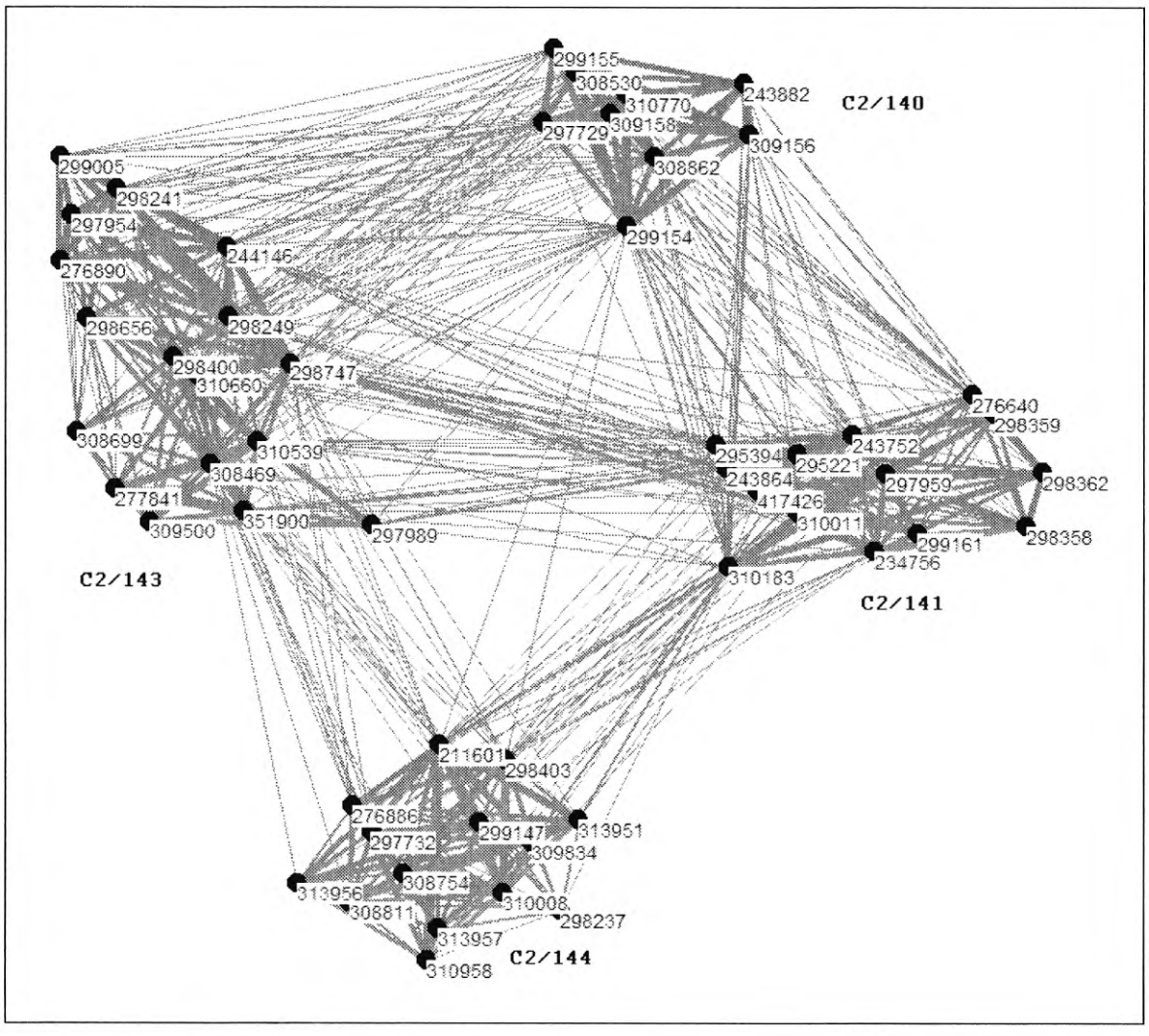
Table 5-25: The Frequency Distribution of Journal Subject Categories in Cluster C3/27

Frequency	Journal Subject Categories
19	Optics
19	Physics, Atomic, Molecular & Chemical
16	Physics, Multidisciplinary
3	Physics, Condensed matter
1	Multidisciplinary sciences
1	Physics, Applied

In the graph depicting C3/27, each of the four C2 clusters is clearly demarcated (see Figure 5-39). The density *D* is 0.43 and the AvgCS(C) 1.97, hence both values are clearly below the average.

[52] Bose-Einstein condensation is the collapse of atoms into a single quantum state.
[53] The exact number of contributing disciplines will not be given by assigned journal subject categories as these are journal classifications, covering the scope of journals, not the scope of the individual article.

Figure 5-39: The Configuration of Core Documents in Cluster C3/27



Note: Kruskal's stress is 0.08.

The field expert presented in this case an elaborated evaluation where not only misplaced core documents on the C1-level were considered, but also minor deviations between their research foci.

The remarks given by the field expert with regard to the cluster composition at the C1 level are as follows:

i. In cluster C2/140/ C1/367, core document 308862 caused some uncertainty as the subject content as reflected by its title was not completely transparent.

ii. In cluster C2/140/C1/155, core document 299154 had a somewhat deviating focus in comparison with other cluster members.

iii. In cluster C2/143/C1/353, core document 298656 had a slightly deviating focus in comparison with other cluster members. Also, core document 308469 and core document 308699 cohered, but were

considered somewhat deviating in relation to core document 277841 and core document 308699, which formed a coherent pair. Hence, this cluster "sprawled" slightly in terms of cluster coherence.

iv. In cluster C2/143/C1/362, core document 297989 deviated somewhat from the other core documents in C3/27.

v. No core document deviated to the extent that it should be considered as clearly misplaced.

vi. Concerning C2-clusters, the field expert's opinion was that all constituent C1-clusters shared the same research focus, hence, all C2-clusters belonged to the same area of research. Conclusively, some deviations on the C1 level were detected and when core documents were aggregated to higher levels, a common research theme for all core documents in C3/27 is seen.

4.4.4 Cluster C3/29: "Carbon-Nano-Tubes"

This cluster contained 22 core documents distributed over two C2-clusters and five C1-clusters:

C2/27: C1/1018; C1/1416
C2/28: C1/549; C1/1072; C1/1137

Core documents in C3/29 focus on carbon-nano-tubes (CNTs) from different angles.[54] A total of 11 different journal subject categories were assigned the journals in which core documents in this cluster were published (see Table 5-26).[55]

[54] Carbon nano tubes are cylindrical carbon molecules with properties that make them potentially useful in extremely small scale electronic and mechanical applications. They exhibit unusual strength and unique electrical properties, and are efficient conductors of heat.

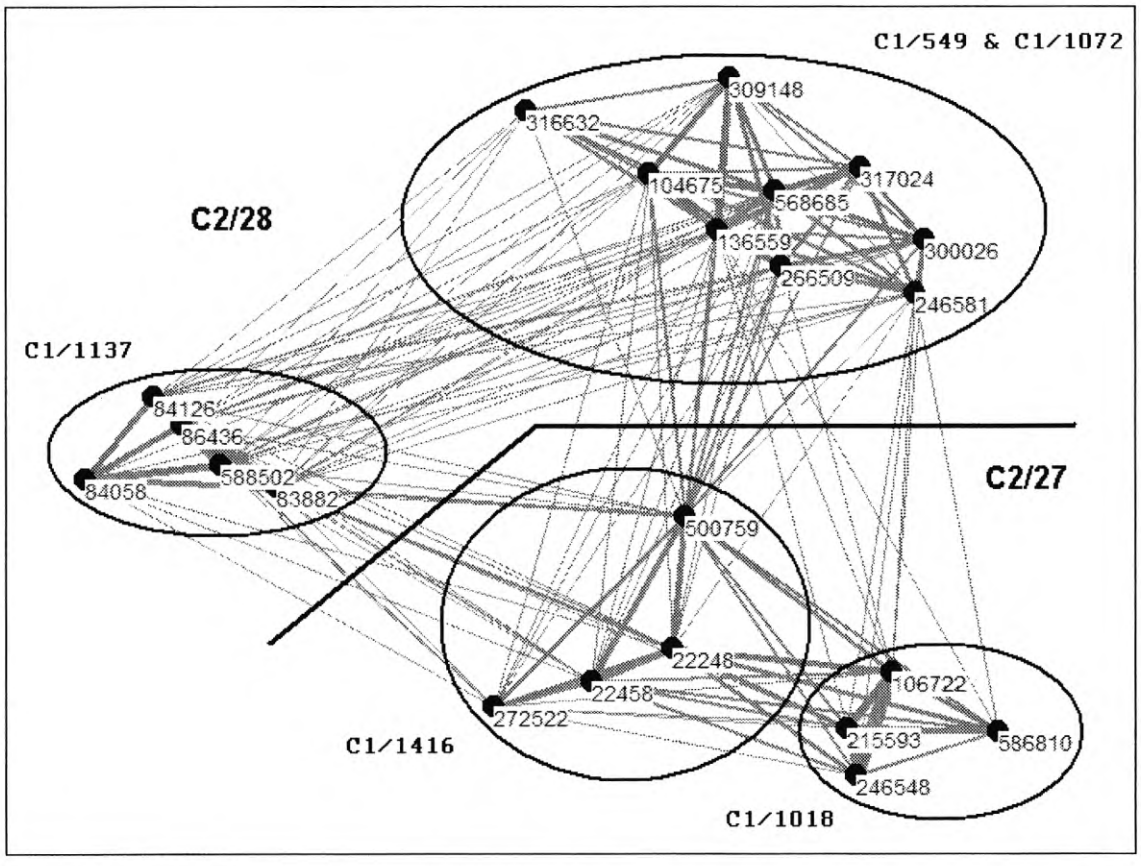
[55] The exact number of contributing disciplines will not be given by assigned journal subject categories as these are journal classifications, covering the scope of journals, not the scope of the individual article.

Table 5-26: The Frequency Distribution of Journal Subject Categories in Cluster C3/27

Frequency	Journal Subject Categories
5	Chemistry, analytical
5	Physics, applied
4	Physics, condensed matter
3	Chemistry, physical
3	Materials Science, multidisciplinary
2	Physics, atomic, molecular & chemical
1	Biochemical Research Methods
1	Engineering, electrical & electronic
1	Multidisciplinary sciences
1	Physics, multidisciplinary
1	Polymer science

In the graph depicting C3/29, a complex and less clear cluster structure is reflected. Hence, the division of the map in C2-clusters and the subdivision in C1-clusters are not clearly mirrored by the configuration of the graph representing C3/29 (see Figure 5-40). The density *D* was 0.70 and the *AvgCS(C)* was 2.45. Hence, the general level of interconnectedness is above the average, but the average strength of links is below the average.

Figure 5-40: The Configuration of Articles in Cluster C3/29



Note: i. The angled line dividing the map indicates the border between cluster C2/27 and cluster C2/28.
ii. Kruskal's stress was 0.07.

The more complicated structure was also reflected in the field expert's evaluation. To begin with, cluster C1/1416 contained one misplaced core document (500759) as did cluster C1/549 (core document 246581).

Moving to the C2 level, in C2/27, both C1/1018 and C1/1416 handled CNT growth, though C1/1018 focused on the growth of aligned CNT on patterned substrates whereas C1/1416 was about non-aligned growth.

In C2/28 (containing C1/549, C1/1072 and C1/1137), C1-clusters focus on CNTs from divergent perspectives with no obvious common theme which would justify their fusion to a C2 cluster.

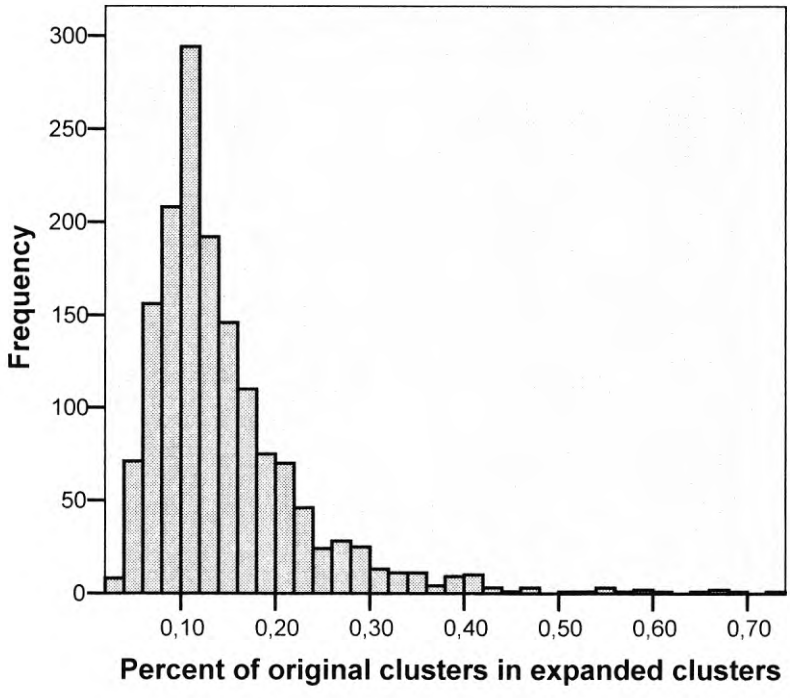
Concerning the subject relationship between cluster C2/27 and cluster C2/28, all C1-clusters explicitly focused on CNTs except for C1/1137 where the interest in CNTs was deemed secondary. [56] However, the field expert renounced the evaluation of the relevance of merging the C2-clusters.

[56] Cluster C2/1137 focused primarily on film-electrodes though all but one core document title had the term "carbon nanotubes" in the title.

4.5 The Expansion of C1-Clusters

In order to examine the extent to which the proposed method gives rise to a fragmentation of research specialties when applied for core document mapping, a complete mapping of significant links connecting core documents in a C1-cluster with core documents extrinsic to the C1-cluster is needed. Computing all such links with a NCS $\geq$ 0.25, the ability of C1 clusters to expand was assessed. In this experiment, the expansion of clusters was tried on all clusters with a size > 1.[57] It was found that on the average, a cluster could expand by eight times its size, consequently 12.5 percent of the articles in an expanded cluster typically constituted the original cluster[58] (see Figure 5-41).

Figure 5-41: The Distribution Of Shares Of Original Clusters In Expanded Clusters



As can be seen from Figure 5-49, only a few core document clusters constitute 50 percent or more of the expanded clusters. The median of the distribution is 0.125, in line with the mentioned factor 8 when calculating the size of an original cluster.

The correlation between original cluster size and share of original articles in an expanded cluster showed a positive correlation with a value of *r* of +0.64. This means that articles in larger original clusters to a lesser extent were associated with articles extrinsic to the original cluster.

---

[57] This means links between core documents in C1-clusters with a size > 1 and all other 5,771 core documents from the first level of cluster fusion.

[58] Let N be the number of documents in the original cluster. Let T be the number of documents in the expanded cluster: N ` 8= T, N/T= 0.125 and 8 ` 0.125= 1 (100%).

In Table 5-27, the poles of clusters with regard to shares of original clusters in expanded clusters are displayed by reversed sort orders, giving the range. The first 30 from each direction are shown.

Table 5-27: The Expansion of C1-Clusters

A1	B1	C1	D1	E1	A2	B2	C2	D2	E2
1060	5	249	174	3%	75	11	34	4	73%
1395	3	120	97	3%	841	9	23	4	69%
150	2	84	63	3%	159	22	40	11	67%
1445	4	172	108	4%	1381	8	29	4	67%
306	2	64	53	4%	1313	11	14	6	65%
169	2	59	52	4%	798	8	33	5	62%
458	2	85	52	4%	1527	9	31	6	60%
223	2	57	50	4%	407	7	29	5	58%
146	2	51	47	4%	1522	11	31	8	58%
543	2	57	47	4%	919	10	45	8	56%
86	2	58	46	4%	457	11	24	9	55%
1338	3	81	68	4%	931	13	70	11	54%
156	6	334	135	4%	1528	9	31	8	53%
186	3	85	65	4%	1309	9	37	9	50%
242	2	46	43	4%	1394	12	95	13	48%
987	2	50	42	5%	432	11	58	12	48%
489	3	96	63	5%	720	8	32	9	47%
378	2	43	41	5%	1462	10	48	12	45%
1157	4	155	82	5%	825	9	45	12	43%
331	3	120	61	5%	859	8	29	11	42%
54	2	42	39	5%	1531	8	39	11	42%
1155	4	180	77	5%	736	20	80	28	42%
1234	6	419	115	5%	1224	9	51	13	41%
301	2	44	38	5%	95	11	53	16	41%
372	2	47	38	5%	1351	11	53	16	41%
1154	3	106	57	5%	658	8	46	12	40%
1374	4	190	75	5%	807	6	37	9	40%
786	3	94	56	5%	1004	6	39	9	40%
318	2	42	37	5%	1314	6	42	9	40%
16	2	53	37	5%	806	6	46	9	40%

Note: i. Columns A1/A2 hold the cluster identity numbers.
ii. Columns B1/B2 hold the sizes of the original C1 clusters.
iii. Columns C1/C2 hold the numbers of links extrinsic to C1 clusters.
iv. Columns D1/D2 hold the numbers of added articles to C1 clusters.
v. Columns E1/E2 hold the shares of original articles in the expanded clusters.
vi. The first five columns from the left, A1 to E1, are sorted in ascending order based on column E1 and the next five columns, A2 to E2, are sorted in descending order based on column E2.

Conclusively, it has been shown that a large number of core documents can be added to core document clusters on the C1-level of fusion by tracking strong links between core documents. This is in line with the suggestion that the mapping of articles linked to core documents would facilitate the coverage of whole research fronts (Glänzel & Czerwon, 1995; 1996). In this case, only

strong links to other core documents were applied and a strong subject relationship between core document clusters and the added core documents could be presumed. The examination of a few samples of such expanded clusters did not contradict this presumption. As an example, cluster C1/ 203 was expanded. Originally this cluster was composed of three articles, all on bio-rhythms. Articles are presented by article number, title, journal title and journal subject category as follows:

i. 321110/ Light and Circadian Regulation in the Expression of Lhy and Lhcb Genes in Phaseolus-Vulgaris/ *Plant Molecular Biology*/ **biochemistry & molecular biology**; **plant sciences**

ii. 321536/ The Circadian Clock - A Plants Best Friend in a Spinning World/ *Plant Physiology*/ **plant sciences**

iii. 401249/ Light-Regulated Translation Mediates Gated Induction of the Arabidopsis Clock Protein Lhy/ *EMBO Journal*/ **biochemistry & molecular biology**

In this cluster, between 14 and 16 common references connect the bibliographically coupled pairs of papers and a total of ten references are common to all papers (the total number of references for a pair in brackets). They are:

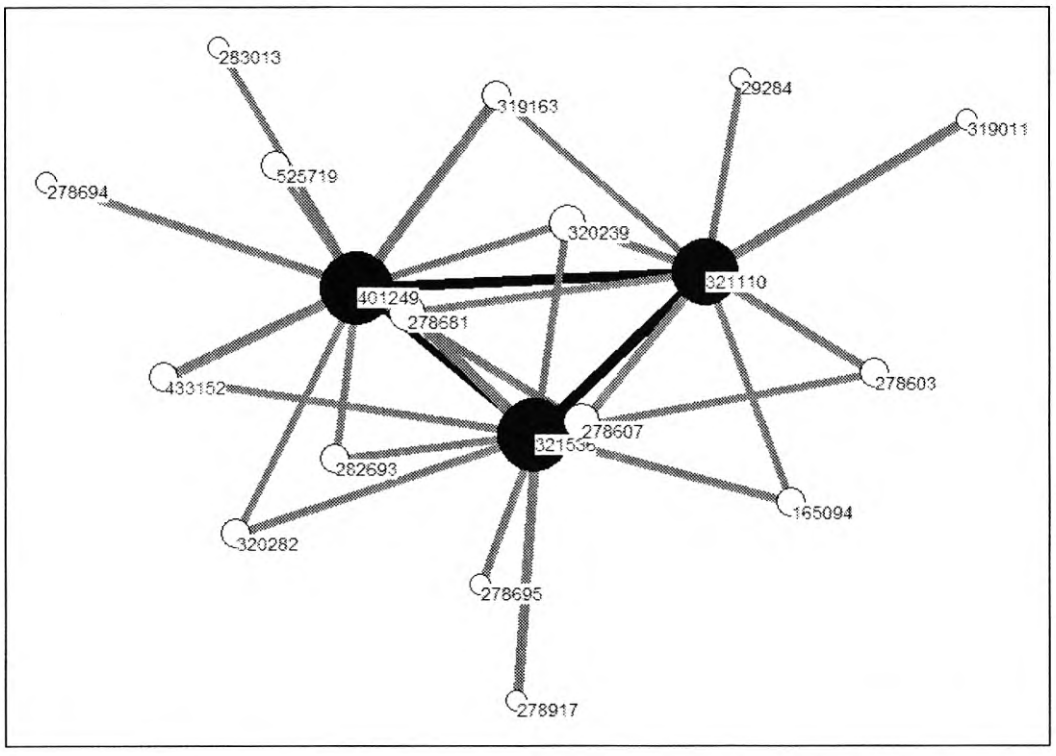
15 (88) 321110-321536
16 (88) 321110-401249
14 (99) 321536-401249

This cluster is linked to 16 other papers extrinsic to the cluster with a *NCS* of at least 0.25 through a total of 28 links as follows:

8/ 321110
10/ 321536
10/ 401249

Expanding the cluster on basis of these links, the cluster could be depicted as an incomplete graph, where the density, *D*, is decreased to 0.18 from the default value of 0.1 (see Figure 5-42).

Figure 5-42: The Expansion of C1/ 203 Depicted by MDS



Note: Cluster C1/203 expanded with 16 unique links to an incomplete graph of 31 edges and 19 vertices. Sizes of circles representing clusters are proportional to the number of links in which a core document occurred (in the expanded cluster) and the width of connecting lines to the *NCS*. Darker lines connecting darker circles depict the original complete subgraph (cluster C1/203). *D* for the incomplete graph (with regard to the applied threshold of NCS) C1/203 was 0.18 and Kruskal's stress 0.06.

The titles and journal subject categories of the added core documents are presented in Table 5-28.

Table 5-28: The 16 Core Documents by which Cluster 203 was Expanded

29284/ Surface-Plasmon Resonance Spectroscopy (Spr) Interaction Studies of the Circadian-Controlled Tomato Lhca4-Asterisk-1 (Cab-11) Protein with Its Promoter/ **biology; physiology**
165094/ Suite of Photoreceptors Entrains the Plant Circadian Clock/ **biochemistry & molecular biology; plant sciences; cell biology**
278603/ Arabidopsis Pseudo-Response-Regulator7 Is a Signaling Intermediate in Phytochrome-Regulated Seedling Deetiolation and Phasing of the Circadian Clock/ **biochemistry & molecular biology; plant sciences; cell biology**
278607/ The Time-for-Coffee Gene Maintains the Amplitude and Timing of Arabidopsis Circadian Clocks/ **biochemistry & molecular biology; plant sciences; cell biology**
278681/ Comparative Genetic-Studies on the Aprr5 and Aprr7 Genes Belonging to the Aprr1/Toc1 Quintet Implicated in Circadian- Rhythm, Control of Flowering Time, and Early Photomorphogenesis/ **plant sciences; cell biology**
278694/ The Evolutionarily Conserved Osprr Quintet - Rice Pseudo- Response Regulators Implicated in Circadian-Rhythm/ **plant sciences; cell biology**
278695/ Characterization of the Aprr9 Pseudo-Response Regulator Belonging to the Aprr1/Toc1 Quintet in Arabidopsis-Thaliana/ **plant sciences; cell biology**
278917/ Response Regulator Homologs Have Complementary, Light- Dependent Functions in the Arabidopsis Circadian Clock/ **plant sciences**
282693/ 2 Arabidopsis Circadian Oscillators Can Be Distinguished by Differential Temperature Sensitivity/ **multidisciplinary sciences**
283013/ Circadian Phase-Specific Degradation of the F-Box Protein Ztl Is Mediated by the Proteasome/ **multidisciplinary sciences**
319011/ The Novel Myb Protein Early-Phytochrome-Responsive1 Is a Component of a Slave Circadian Oscillator in Arabidopsis/ **biochemistry & molecular biology; plant sciences; cell biology**
319163/ Dual Role of Toc1 in the Control of Circadian and Photomorphogenic Responses in Arabidopsis/ **biochemistry & molecular biology; plant sciences; cell biology**
320239/ A Link Between Circadian-Controlled Bhlh Factors and the Aprr1/Toc1 Quintet in Arabidopsis-Thaliana/ **plant sciences; cell biology**
320282/ Cell Autonomous Circadian Waves of the Aprr1/Toc1 Quintet in an Established Cell-Line of Arabidopsis-Thaliana/ **plant sciences; cell biology**
433152/ The Arabidopsis-Srr1 Gene Mediates Phyb Signaling and Is Required for Normal Circadian Clock Function/ **developmental biology; genetics & heredity**
525719/ Fkf1 Is Essential for Photoperiodic-Specific Light Signaling in Arabidopsis/ **multidisciplinary sciences**

Note: Core documents are presented with article numbers, article titles and journal subject categories. Subject categories are in extra bold style.

As can be seen from Table 5-28, all articles connect to the original research focus of cluster C1/203.

4.6 Summary

Applying the complete link cluster method on core document data resulted in a first partition where the majority of clusters were relatively small and the median cluster size was 4 for the selected set of clusters with a size ≥ 3. On each fusion level, a share of clusters that did not fulfill the requirements for cluster fusion emerged. This way, by each level of cluster fusion, C1 to C3, the original set of core documents was reduced as the sizes of clusters increased. The stepwise loss of core documents and simultaneous increase in cluster size is presented in Table 5-29.

Table 5-29: Three Levels of Cluster Fusion: Effects on Document Populations, Frequency of Clusters and Cluster Sizes

Level of Fusion	No. of Clustered Core Documents	No. of Clusters	Median Cluster Size
C1	4,477	1,000	4
C2	3,524	212	24
C3	1,763	38	37

Note: i. The calculation of median cluster size does not include singleton clusters.
ii. On the C1 level, clusters have a minimal size of three articles.
iii. On the C2- and C3- levels, clusters are composed by at least two objects (clusters from earlier fusion levels).

Concerning the aspect of external cluster isolation, by each level of fusion, the share of isolated clusters was increased while the strength of association between clusters was weakened. At the same time, the internal cluster coherence was weakened too. Comparing levels, the most drastic change with regard to the separation between clusters take place when moving up to the C2 level, while the most drastic change with regard to cluster coherence takes place when moving up to the C3-level (see Table 5-30).

Table 5-30: Three Levels of Cluster Fusion: Effects on Cluster Coherence and Cluster Isolation

Level of Fusion	Percentage Isolated Clusters	Median AvgCS(C, C')	Mean AvgCS(C)	Median *D*
C1	5	0.65	10.58	1.00
C2	11	0.02	7.95	1.00
C3	16	0.00	3.64	0.64

With regard to statistical data, the optimal level of cluster fusion should be the C2-level. The reasons are as follows:

i. On the C1-level, associations between core documents in different clusters are strong.

ii. On the C2-level, clusters are generally still coherent and well separated.

iii. On the C3-level, clusters are considerably less coherent.

These findings should be related to the field experts' evaluations of the four C3-clusters. On the C1-level, clusters were generally considered subject coherent. On the C2 level, in one case (C3/29), one C2-cluster was considered artificial. On the C3-level, only two of four C3-clusters could be evaluated with regard to the merging of C2-clusters, and one of these was considered irrelevant. Hence, field experts' evaluations did not contradict statistical findings.

Lastly, though findings regarding iterated clustering of core documents indicated the breaking up of specialties by the generation of coherent C2-clusters, they did not cover for all associations between core documents in a C1-cluster and core documents extrinsic to it as the partition in clusters itself breaks up links. Hence, mapping all links between core documents in a cluster and core documents extrinsic to the cluster with a minimal NCS of 0.25, it was clearly shown that core document clusters on the C1-level constitute fragments of larger research themes.

CHAPTER 6: DISCUSSION AND CONCLUSIONS

This chapter wraps up the study undertaken. It begins with a discussion of the empirical findings which are summarized and discussed. The last section gives the conclusions drawn on this study.

1. DISCUSSION

1.1 Cases 1 to 3

1.1.1 The Relevance of Clusters Generated by the Complete Link Cluster Method

A small research field would generally imply weaker and fewer links of bibliographic coupling as a result of a lower publication output and a smaller base literature. Hence, methods of bibliographic coupling would generally be less applicable on research fields with a lower publication output (Glänzel & Czerwon, 1996). Furthermore, methods of bibliographic coupling should be applied with quite severe thresholds of NCS in order to secure significant associations between articles (Sen & Gan, 1983; Glänzel & Czerwon, 1996). However, with regard to Case 1 and Case 3, it was shown that the proposed method is capable of generating relevant clusters also on low levels of NCS (see Sub-section 3.1 in Chapter 4).

In Case 2, the larger population of articles from the field of Organic Chemistry was applied as the test arena, facilitating the application of considerably more severe thresholds, and also the identification of a delimited set of core documents. This was also reflected by higher values of AvgCS(C) and in a comparably lower share of misplaced articles (2 percent). The microanalysis in Case 2 illustrated the ability of the method when applied on a single but large research field to map core documents and generate relevant clusters as no article was regarded as misplaced.[59] The high relevance of generated clusters in the microanalysis of core documents was underlined by the agreement between results accomplished by the complete link clustering and MDS.

1.1.2 The Extent and Nature of Deviations Between Results Generated by the Complete Link Cluster Method and Results Generated by Intellectual-Manual Partitions

It has been shown that in all cases there were large differences between the intellectual-manual clusterings performed by the field experts and the complete link clusterings. To begin with, the distribution of articles over clusters deviated in the sense that the partitions generated by the complete link cluster method resulted in more and smaller clusters and a lesser concentration of articles to clusters. This difference was explicitly illustrated by Pratt's measure of concentration (see Table 6-1).

[59] One article was not evaluated though.

Table 6-1: The Concentration of Articles to Clusters in Cases 1 to 3

Case	Complete Link Clusters		Experts' Clusters	
	No. Of Clusters	Pratt's Measure	No. Of Clusters	Pratt's Measure
1	17	0.13	10	0.39
2	44	0.17	17	0.57
3	40	0.07	36	0.42

With regard to internal coherence, the field experts' clusters were generally less coherent with respect to both the strength and the density of links (see Table 6-2).

Table 6-2: The Internal Coherence of Clusters in Case 1 to 3

Case	Complete Link Clusters	Experts' Clusters	
	Md AvgCS(C)	Md AvgCS(C)	*D*
1	3.67	1.90	0.61
2	13.94	3.91	0.30
3	4.17	1.65	0.33

Note: i. Clusters generated by the complete link cluster method have a default maximal value of *D* (1.0).
ii. Singleton clusters generated by the field experts are excluded in the calculations.

With regard to the aspect of external isolation, clusters generated by field experts were generally less isolated as reflected by the shares of all 2-combinations of clusters that were coupled. With regard to the AvgCS(C, C'), but with the exception for Case 1, the association between clusters were weaker when the complete link cluster method was applied. However, differences were not pronounced (see Table 6-3). [60]

[60] The complex interplay between measures of coherence and isolation must be interpreted on the level of a particular case. This may provide the researcher with a detailed understanding of a field's cognitive structure and the impact of chosen methods. Here, the aim of analysis is delimited to study deviations between partitions generated by two different methods.

Table 6-3: The External Isolation of Clusters in Case 1 to 3

Case	Complete Link Clusters			Experts' Clusters		
	A	B	C	A	B	C
1	0.19	0.42	0	0.16	0.78	0
2	0.24	0.12	5	0.40	0.35	4
3	0.17	0.07	5	0.36	0.13	0

Note: A-columns contain median coefficients of AvgCS(C, C'). The median was calculated excluding any isolated clusters that occurred; B-columns contain the shares of 2-combinations of clusters that were coupled and C-columns contain the number of isolated clusters.

On a detailed level, the agreements between partitions were assessed by tabulating distributions by different sort orders, facilitating an exact visualization of deviations. Generally, little agreement between partitions was seen and some extreme deviations were found. For all, partitions generated by the complete link cluster method resulted in a more fine graded division, and often in a split up of experts' clusters, which was also illustrated by MDS.

1.1.3 A Commentary on and Comparison of Methods of Partition

Contrasting the relatively high degree of relevance in clusters generated by the complete link cluster method with the pronounced deviations between the partitions generated by this cluster method and experts' partitions, one may assume that there exist alternative classifications for Cases 1 to 3. Below are some suggestions that are in line with this assumption.

The complete link cluster method applies common references as a measure of similarity exclusively. The intellectual clustering, on the other hand, was foremost based on semantic relations between titles (and abstracts) in different articles and largely independent of common references. Semantic relations between articles may well exist also when citation relations are weak or absent, as may citation relations between articles exist when there is an unclear or absent semantic relation. The latter would, for instance, be the case when different specialties are merged into new interdisciplinary areas and different terms are used to denote the same objects or phenomena. Hence, the two methods of partition should generate similar results only if semantic relations and citation relations converge.

Moreover, the classification of articles accomplished by the complete link cluster method is dependent on how current research proceeds and undergoes changes as reflected by the use (referencing) of previous research. Due to the dynamic aspect of research, classifications (clusters) are not easy to anticipate. An associated issue is the demarcation of research themes with regard to the choice of hierarchical level for a cluster solution. As there exists no common framework for the demarcation of a field's division in specialties, the delineation of borders between specialties or disciplines may well provide

difficulties.[61] These aspects were to some extent reflected in some of the field experts' comments, which indicated that more than one cluster solution may be acceptable. In Case 1, the field expert admitted the split up of some expert clusters when compared with the clusters generated by the complete link cluster method. In Case 2, the field expert noted that a new set of principles, hard to anticipate on beforehand, for partitioning and classification emerged when studying clusters generated by the complete link cluster method. In Case 3, the field expert concluded that the classification of articles could have different points of departure, and that a more fine graded partition of articles as well as the merging of some groups may be equally valid.

1.1.4 The Effects of Threshold Settings and Method of Partition on the Original Populations of Research Articles

As discussed in Sub-section 3.3 in Chapter 2, the issue of the extent to which topics covered by a population of research articles are identified by the applied method should be of interest. In this study the "recall" of relevant articles can not be directly assessed as the number of articles that have a cognitive (semantic) relation to a certain cluster is not known. However, assessing the effects of applied thresholds and cluster method on the sizes of the original populations of research articles, a coarse estimate of the proposed method's ability to intercept current research themes of the populations under study could be provided.

As selection criteria always will lead to a reduction of a population of articles, the diminishing of the sizes of the original populations of articles is granted. Several factors affect the extent to which an original population will be diminished. The more important factors are:

i. the extent of consensual referencing of the field under investigation;

ii. the set threshold of coupling strength or NCS ; and

iii. the choice of cluster method.

Concerning (i), this should be the most important factor deciding the extent to which research themes of the original population of articles is covered. Hence, when similar topics are treated but the referencing is less consensual (or attentive), articles will be lost. The share of articles not belonging to any cluster of the set minimal size is thus a reflection of the extent of non-consensual referencing for a particular population.

With regard to (ii), there exists no straight forward method for deciding the most appropriate threshold, hence empirical experience would guide decisions of thresholds and methods may initially be more or less arbitrary (cf. Sub-section 3.3 in Chapter 2). Generally, a large research field where specialties

[61] As stated in Tijssen (1992, p. 31), "[u]nlike geographical maps, maps of science are not directly related to the physical world" as there exists no common frame of reference.

have a clear and consensual focus would provide stronger links of bibliographic coupling and more choices of threshold setting.

With regard to (iii), the choice of cluster method has an impact on the sizes of clusters. Generally, the more severe conditions to fulfill for the merging of articles to clusters, the greater the number of smaller sized clusters and the subsequent loss of articles when excluding clusters below a stipulated threshold of cluster size.

It is clear that points (i) to (iii) are interrelated and that this interrelationship is complex and difficult to foresee. The empirical findings in Cases 1 to 3, reflect the impact of these factors on the original sizes of populations, which is illustrated in Table 6-4.

Table 6-4. The Successive Diminishing of the Original Populations.

Case	Original Size	A	B
1	232	185	63
2	14,389	268	183
3	879	579	130

Note: Column A shows the sizes of the populations after threshold setting of the coupling Strength or NCS (Case 2) and column B shows the sizes of the populations when clusters containing less than three articles have been excluded.

Starting with Case 1, the original size of the population of articles was 232. The method of noise reduction (the application of a threshold of one coupling unit) implied a further diminishing of the set by approximately 20 percent to a total of 185 articles. After clustering and exclusion of clusters where the size < 3, only 34 percent of these articles remained. Conclusively, the total reduction was 73 percent.

In Case 2, the filtering out of bibliographically coupled pairs with a NCS below 0.25 from the remaining articles and applying a threshold of four links at the same threshold of NCS brought about a reduction of approximately 98 percent. Considering the effect of applying cluster size threshold, the total reduction of articles was 99 percent.

In Case 3, with regard to the noise reducing actions taken, the same approach of threshold setting as in Case 1 was applied (one coupling unit). This brought about a diminishing of the set of articles by 34 percent to a final set of 579 articles. This set was further reduced by 78 percent to 130 articles by excluding clusters containing less than three articles. Hence a total reduction by 85 percent.

It can be concluded that even if low thresholds (Cases 1 and 3) of coupling strength are applied, a notable reduction of the sizes of the original populations of articles takes place.

1.1.5 Implications of Findings

Only a small fraction of articles in the original populations was included in the mappings, most likely implying the absence of research themes as well as articles potentially relevant for the subject foci of clusters. Generally, there exists a clash between relevance and interception as severe thresholds imply a considerable loss of articles, whereas the absence or application of low thresholds may impair the relevance of clusters. A similar clash should exist with regard to the choice of cluster method as a method with severe conditions to fulfill (e.g. the complete link cluster method) would bring about an increased loss of articles in comparison with a more generous method (e.g. the single link cluster method), but promote relevance.

When the prime objective is to find relevant information, a gearing of thresholds may tentatively be applied for information provision purposes. As was shown, science fields of different sizes and referencing characters may be mapped by the applied method and when the network of bibliographically coupled articles of a selected population so allows for, thresholds may be varied so that both cores of consensual research as well as more comprehensive but perhaps less significant cluster structures are identified. With regard to the latter, most probably, the application of lower thresholds in Case 2 would still generate useful but perhaps less lucid information.

The fact that partitions generated by field experts consequently and strongly deviated from partitions generated by the complete link cluster method means that the proposed method did not converge with field experts' comprehensions of fields' cognitive structures. Hence, if accepting field experts' apprehensions of scientific structures as valid points of reference, the findings indicated that the proposed method may not clearly identify conceptualized structures, hence its capability of laying out the cognitive structures of specialties should not only on theoretical grounds (see Sub-section 3.2 in Chapter 2) be ambiguous. However, subject coherent clusters containing relevant information were generated over all three cases and the method was capable of identifying smaller, coherent research foci on comparably low levels of NCS. The meaning of the deviations between field experts' apprehensions of cognitive structures and structures generated by the applied method should provide incentives for further research. It is clear, however, that the deviations were not only about more or less fine graded partitions, but also signaled a difference of how research concepts are associated.

The variations between fields with regard to the estimated relevance of clusters should be commented. The considerably much stronger links of NCS arrived at in Case 2 (Md. NCS=0.31) were reflected by a higher relevance of clusters (2 percent misplaced articles) in comparison with Cases 1 and 3. The difference between Cases 1 and Case 3 with regard to the relevance of clusters (7 percent misplaced articles in Case 1 vs. 13 percent misplaced articles in Case 3) could only tentatively be assigned to the difference of the median NCS as this difference was not pronounced (0.15 in Case 1 and 0.09 in Case 3). The comparably less severe relative diminishing of the population in Case

2 specifically due to the clustering process (see Table 6-4) should preliminary be explained by the applied thresholds.

1.2 Case 4

The point of departure in the following discussion is in the final set of core documents containing 4,477 articles. This set was accomplished by a gradual reduction of the original set of 6,060 core documents by 26 percent when thresholds of NCS and cluster size were applied.

1.2.1 The Extent of Fragmentation Imposed by the Applied Method

It was shown that the applied method leads to a fragmentation of research themes. On the average, a core document cluster could increase its size by a factor of eight and only a few clusters were expanded by less than half their sizes. The effect of fragmentation was illustrated by example where it also was shown that the adding of core documents brings about a decrease of cluster coherence, measured as *D*. Hence, the expansion of clusters is at a cost of a presumably diminished relevance.

1.2.2 The Impact of Iterated Clustering on the Overall Cluster Structure

With the starting point in a large set of smaller clusters, the fusion of clusters at two subsequent levels showed an increasing loss of core documents as larger aggregations of core documents were formed. This loss was due to the generation of singleton clusters and isolated clusters emerging at each level of cluster fusion. At each subsequent level of cluster fusion, the general tendency was that by the increase of cluster size, there was a simultaneous decrease of cluster coherence and an increase of the external isolation of clusters. This means that increasingly less relevant clusters were formed but also that clusters got more isolated.

1.2.3 The Optimal Level of Cluster Fusion

It was found that the second level of cluster fusion (C2) should be the optimal level. This could be the concluded on the following grounds:

i. On the first level of cluster fusion (C1), a large share of clusters were associated with other clusters through relatively strong links.

ii. On the second level of cluster fusion, the internal coherence remained strong and at the same time clusters were generally more isolated.

iii. On the last level of cluster fusion, the drop of cluster coherence was considerable, indicating the generation of more subject inconsistent clusters.

Field experts' evaluations did not contradict these findings.

1.2.4 Implication of Findings

Though empirical findings speak in favor for the second level of cluster fusion as the most appropriate, it was shown that a few clusters on the C1-level are nearly complete in terms of extrinsic associations and that a few clusters on the C3-level may be relevant. The example of cluster fusion over three levels illustrated that the association between disciplines through their research foci may provide interesting links which may give an overview of a problem area. The proposed method is, however, not likely to be applicable on the last level of cluster fusion (C3). Moreover, the quite severe loss of core documents generated by iterated clustering would require the interpretation of data from the preceding levels if a more comprehensive mapping should be accomplished. Hence, it is suggested that at least the two first levels of fusion are applied, including singleton clusters and isolated clusters and that mapping results be interpreted from bottom to top (or top bottom) as the cluster merging itself contain important information.

The assessed effects of fragmentation implies that the proposed method when applied for core document clustering do not identify and map research themes exhaustively, but rather smaller cores of referencing consensus. Also, findings showed that approximately a quarter of the final population of core documents where lost when clustered, given the applied minimum size of clusters.

1.3 Reflections on Findings in Relation to Previous Research

Several results connect to previous findings and theoretical considerations in the literature on bibliographic coupling and cocitation cluster analysis. First, claims that the method of bibliographic coupling is capable of associating documents that have a similar research focus (e.g. Vladutz & Cook, 1984; Peters, Braam & van Raan, 1995) was confirmed by the relatively high degree of relevance in clusters generated by the proposed method. The application of the complete link method, in line with coupling criterion B, originally suggested by Kessler (1960) and the suggestion of "cliques" as one type of bibliographically coupled document groups (Sen & Gan, 1983), resulted in small but compact and generally subject consistent clusters. Hence, the problems of "chaining" encountered in cocitation cluster analysis (cf. Griffith, Small, Stonehill & Dey, 1974) was avoided. The effect of fragmentation, or more precisely, the split up of research themes in smaller clusters, also encountered in cocitation cluster analysis (Braam, Moed & van Raan, 1991), was conspicuous. The issue of fragmentation is also related to the setting of thresholds of coupling strength. These problems was approached by Small and co-workers (Small & Sweeney, 1985) by implementing variable level clustering in order to find the best cluster solution. The problems of threshold setting was avoided in the case of core document mapping, as previous empirical findings would guide the setting of these (cf. Glänzel & Czerwon, 1995; 1996).

The effect of the dependency of consensual referencing and the associated problems of threshold setting was observed in this study and could be

recognized as a severe diminishing of document populations. This type of problem has also been recognized in research in cocitation clustering where findings have shown that only parts of document populations relevant to identified research topics were revealed (cf. Braam, Moed & van Raan, 1991). This concerns the issue of the exhaustiveness of citation based science mapping. Braam, Moed and van Raan recognized that this question demands a comparison of cluster solutions on different levels of thresholds and the simultaneous use of complementary methods (ibid.).

Through the criticism of the cocitation cluster analytical method, the statistical instability of the method (Oberski, 1988) and inconsistent results (Leydesdorff, 1987), the much varying results possible to arrive at by just tampering one of several affecting variables were highlighted. This connects to the difficulty to *empirically* arrive at method applications of citation based mapping that may generate optimal results, which could be illustrated as follows. Assume that the three more important variables are selected for empirical testing. Let these variables be the following ones:

i. population;

ii. choice of cluster method; and

iii. threshold of coupling strength.

Next, to each of these variables is assigned three sub-variables. The number of research settings required should then be 27. It would also be reasonable to include other multivariate techniques, e.g. factor analytical approaches, which should increase the number of research settings further.

The difficulty to *theoretically* contribute to successful applications of citation based mapping is the absence of a conceptual framework. The significance of this problem was stressed by field experts' comments concerning alternative (and equally valid) mapping solutions and by the fact that the proposed method generated generally relevant clusters much deviating from field experts' clusters. Conclusively, a general problem of citation based science mapping is the absence of a common frame of reference. Therefore, more axiomatic approaches may pay off.

2. CONCLUSIONS

In the study undertaken, a method was suggested for science mapping purposes and evaluated. The suggestion of this method was motivated by the fact that the prevailing method of citation based science document mapping, the cocitation cluster analytical method, can not map the most current published research, a feature that is a characteristic of the proposed method. The cocitation cluster analytical method, on the other hand, is based on a theory which claims that the more central research questions of a specialty can be identified through highly cocited documents. On this ground, it is presumed that the identification of the cognitive structures of specialties may be mapped. This is a feature that could not (as for now) be assigned to the proposed method. It was therefore assumed that none of these methods could substitute each other and that they would be complementary.

Previous research has stated the capability of the bibliographic coupling method to associate subject similar documents with one another and its applicability for IR purposes. However, there is an explicit lack of empirical experience concerning the application of bibliographic coupling in the context of science mapping. Therefore, empirical experience from cluster analytical research in the context of science mapping could only be obtained from the research in cocitation cluster analysis. Based on criticism of the cocitation cluster analytical method and on reported empirical experiments, the following problems were presumed to be of importance also for the application of the proposed method:

i. The dependency of consensual referencing implies that only minor shares of original document populations will be available for analysis.

ii. The lack of a method for the decision of appropriate thresholds of coupling strength implies arbitrary threshold settings.

iii. The choice of the single link cluster method has shown the undesirable effect of chaining (prolonged and loosely bound clusters) and the subsequent generation of macro clusters.

iv. The partition of document populations has brought about the split up of research specialties, an effect of fragmentation of research fields.

Findings confirmed the relevance of each of the above points with regard to the proposed method. These issues may be regarded as general for citation based science mapping applying documents as the analyzed unit. With regard to (i), only a minor fraction of the original populations were available for analysis and the stepwise diminishing of document populations was due to: (1) the filtering out of articles lacking bibliographic coupling relations; (2) the setting of thresholds of coupling strength and (3) the partition of document populations by the applied cluster method in combination with a set minimal cluster size. These three causes of reduction of populations all reflected the impact of and dependency on consensual referencing.

With regard to (ii), no valid method concerning the setting of thresholds of coupling strength was arrived at during the experimental phase. For the three first cases (single field level) considerations were taken with regard to the publication output of corresponding fields, and a 'rule of thumb' approach was applied. Findings showed that the application of severe thresholds of coupling strength implies more relevant clusters, but also that lower levels of coupling strength, necessary to apply in research settings where smaller or younger fields are mapped, are feasible. With regard to the fourth large and multidisciplinary research setting, the specific objective of mapping core documents implied the application of strict rules for the setting of thresholds.

With regard to (iii), the design of the proposed method concerning the choice of cluster method was mainly based on theoretical considerations derived from the statistical literature on cluster analysis and the reported problems and criticism of the use of the single link cluster method in cocitation cluster analysis. Findings showed that the choice of the complete link cluster method resulted in coherent and generally subject consistent clusters. The well known drawback of the single link cluster method was hence steered clear of. However, as mentioned, the strict rules of merging also implied the generation of a large share of smaller sized clusters that from an information provision point of view should be regarded as noise, and the filtering out of these added on to the aforementioned reduction of document populations.

With regard to (iv), the effect of fragmentation was also seen in this study. For the first three cases, this effect was foremost noticed as a decisive difference of number of clusters and cluster sizes between partitions generated by field experts and partitions accomplished by the application of the proposed method. This difference was concluded and summarized applying Pratt's measure of concentration. Concerning the fourth case, the specific properties of core documents and the properties of the proposed method implied a research design aiming at the explicit elaboration of the easily foreseen effect of fragmentation. It was shown that core document clusters to a large extent depicted smaller consensual cores of current published research and that such cores could be expanded considerably also when only strong links were applied. This finding clearly showed that core document clusters constituted minor shares of larger research themes.

It was further illustrated how iterated clustering could connect such cognitively related cores, and the optimal level of iterated clustering was found. The external points of reference accomplished by field experts' evaluations of four examples of iterated clustering did not contradict these findings. Due to the complex relation between fragmentation and relevance (subject consistency) of core document clusters, it could be concluded that the information inherent on all levels of cluster fusion as well as in the process of cluster fusion itself should be used for optimal results.

In the first three research settings relevant clusters were generally generated and it can be concluded that the proposed method has the capability to identify and map current and coherent research themes of a single research field, also when less severe thresholds of coupling strength are applied. The significance

of the information contained in the generated clusters was, however, not unambiguous as it was shown that the proposed method generated clusters strongly deviating from field expert's conceptions of their own fields' structures. This indicates that the proposed method generates information not anticipated by field experts and that this information may have a value of novelty. From another point of view, more congruence between field experts' clusters and clusters generated by the proposed method may have indicated the possibility of replication of expert knowledge and opened up for new lines of research more connected to the elaboration of cognitive and social structures of science.

In the fourth research setting, findings talks in favor for the generation of generally subject consistent clusters on the two first levels of cluster fusion. Findings also indicated that on the third level of cluster fusion, most significant links were exhausted, indicating the upper limit of cluster fusion for the proposed method. It could be assumed that sometimes useful information may be obtained also on this level.

It could be concluded that the proposed method does not apply to traditional mapping objectives, i.e. the elucidation of specialty cognitive structures. Hence, its areas of application should foremost pertain to scientific information provision and be complementary to traditional citation indexing and cocitation cluster analysis.

Further developments of the proposed method in the context of core document mapping and information provision could be accomplished. In particular, it could be suggested that the proposed method could be used as a navigating and information seeking tool. Several applications may be successful and one can be outlined on basis of findings. With a starting point in a complete graph (a core document cluster), the additional expansion by significant links could be used to monitor the radiating associations of articles related to a specific research theme. When additional information of cluster affiliation of such associated articles is added, the navigation in and between scientific structures would be facilitated.[62] The navigation could be geared by varying threshold settings, deciding the maximum radius from each core.

[62] This is actually the basic principle on which the database underlying the empirical study of Case 4 was built. The original idea of using core documents to trace links of associated articles and the subsequent mapping of research fronts was first presented by Glänzel & Czerwon (1995). Hence, the idea presented here is only a modification and expansion of their original idea.

REFERENCES

Ahlgren, P. Jarneving, B. & Rousseau, R. (2003). Requirements for a co-citation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science & Technology*.

Aldenderfer, M.S. & Blashfield, R. K. (1984). Cluster analysis. In Quantitative Applications in the Social Sciences, Vol. 44. London: Sage Publications Inc.

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

Biglan, A. (1973). The difference of subject matter in different academic areas. *Journal of Applied Psychology*. 57(3):195-203.

Braam, R. R., Moed, H. F. & van Raan, A. J. F. (1988). *Mapping of science: Critical Elaboration and new approaches, a case study in agricultural biochemistry*. In International conference on bibliometrics and theoretical aspects of information retrieval. (Eds. Egghe, L. and Rousseau, R.). Amsterdam: Elsevier Science Publishers.

Braam, R. R., Moed, H. F. & van Raan, A. F. J. (1991). Mapping Science by combined Co-citation and word analysis 1: structural aspects. *Journal of the American Society for Information Science*, 42(4):233-251.

Cole, J.R. & Cole, S.C. (1973). *Social Stratification in Science*. Chicago: The University of Chicago Press.

Egghe, L. & Rousseau, R. (1990). *Introduction to informetrics*. Amsterdam: Elsevier.

Everitt, B.S., Landau, S. & Leese, M. (2001). *Cluster analysis*. Fourth edition. London: Arnold.

Fano, R.M. (1956). *Document in action.* New York: Reinhold Publishing Corporation.

Garfield, E. (1979). *Citation Indexing*: *its theory and application in science, technology, and humanities* New York: John Wiley & Sons.

Garfield, E. (1998). From citation indexes to informetrics: Is the tail wagging the dog? *Libri*, 48, (2): 67-80.

Glänzel, W. & Czerwon, H. J. (1995). A new methodological approach to bibliographic coupling and its application to research-front and other core documents, *Proceedings of 5th International Conference on scientometrics and Informetrics*, held in River Forest, Illinois, June 7-10: 167-176.

Glänzel, W. & Czerwon, H. J. (1996). A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. *Scientometrics*, 37(2):195-221.

Griffith, B. Small, H. Stonehill, J. & Dey, S. (1974). The structure of scientific literatures II: toward a macro- and microstructure for science. *Science studies*, 4: 339-365.

Jarneving, B. (2001). The cognitive structure of current cardiovascular research. *Scientometric*, 50 (3): 365-389.

Johnsonbaugh, R. (1997). *Discrete Mathematics*. Prentice Hall International: New Jersey, Forth Edition.

Johnson, D. E. (1998). *Applied multivariate methods for data analysts*. Pacific Grove: Duxbury Press.

Kessler, M. M. (1958). Concerning some problems of intrascience communication. Massachusetts Institute for Technology, Lincoln Laboratory.

Kessler, M. M. (1960). An experimental communication center for scientific and technical information. Massachusetts Institute for Technology, Lincoln Laboratory.

Kessler, M. M. (1962). An experimental study of bibliographic coupling between technical papers. Massachusetts Institute for Technology, Lincoln Laboratory.

Kessler, M.M. (1963a). Bibliographic coupling between scientific papers. *American Documentation*, 14(1):10-25.

Kessler, M.M. (1963b). Bibliographic coupling extended in time: Ten case histories. *Information Storage and Retrieval*, 1:169-187.

Kessler, M.M. (1965). Comparison of the results of bibliographic coupling and analytic subject indexing. *American Documentation*, 16(3):223-233.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrica*, 29 (1):1-27

Kruskal, J. B., Wish, M. B. (1978). *Multidimensional scaling*. In Quantitative Applications in the Social Sciences, vol. 11. London: Sage Publications, Inc.

Leydesdorff, L. (1987). Various methods for the mapping of science. *Scientometrics*. (11):295-324.

Marshakova, I.V. (1973). System of document connections based on references. *Nauchno-Tekhnicheskaya Informatsiya Seriya 2-Informatsionnye Protsessy I Sistemy*, 2(6):3-8.

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.

Martyn, J. (1964). Bibliographic coupling. *Journal of Documentation*, 20:236.

McCain, K. W. (1986). Cocited author mapping as a valid representation of intellectual structure. *Journal of the American Society for Information Science*, 37(3):111-122.

McCain, K.W. (1990). Mapping authors in intellectual space - a technical overview. *Journal of the American Society for Information Science*, 41(6):433-443.

Miller, G.A. (1969). A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology*. 6:169-191.

Mubeen, M.A. (1995). Bibliographic coupling: an empirical study of economics. Annals of Library Science and Documentation, 42(2):41-53.

Noyons, E.C.M. (1999). *Bibliometric mapping as a science policy and research management tool*. Leiden: DSWO Press.

Oberski, J.E.J. (1988). *Some statistical aspects of co-citation cluster analysis and a judgement by physicists*. In Handbook of quantitative studies of science and technology. Ed.:A.F.J.van Raan. Amsterdam: North Holland.

Otte, E. & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*. 28(6):441-453.

Persson, O. (1988). *Measuring Scientific output by online techniques*. In Handbook of quantitative studies of science and technology. Ed.:A.F.J.van Raan. Amsterdam: North Holland.

Persson, O. (1994). The intellectual base and research front of JASIS 1986-1990. *Journal of the American Society for Information Science*. 45(1):31-38.

Peters, H. P. F., Braam, R. R. and 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.

Pratt, A. D. (1977). A measure of class concentration in bibliometrics. *Journal of the American Society for Information Science*. 28, September: 285-292

Pritchard, Alan (1969). Statistical bibliography or bibliometrics? *Journal of Documentation*. 25(4):348-349.

Sen, S. K. & Gan. S. K. (1983). A mathematical extension of the idea of bibliographic coupling and its applications. *Annals of Library Science and Documentation*, 30(2):78-82.

Sharabchiev, Y.T. (1988). Comparative analysis of 2 methods of cluster analysis of bibliographic citation. (In Russian). *Naucno-techniceskaja informacija* / 2.

Sharada, B.A. & Sharma, J.S. (1993). A study of bibliographic coupling in linguistic research. *Annals of Library Science and Documentation*, 40 (4). 25-137.

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(July-August):265-269.

Small, H. (1977). A co-citation model of a scientific specialty: A longitudinal study of collagen research. *Social studies of science*, 7(2):139-166.

Small, H. & Griffith, B. (1974). The structure of scientific literatures I: identifying and graphing specialities. *Science studies*, 4(1):17-40.

Small, H. & Griffith, B. (1983). The structure of the social and behavioral sciences literature. *Stockholm papers in library and information science*. Ed.: Stephan Schwarz. Stockholm: Royal Institute of Technology Library, TRITA-LIB-6021, July.

Small, H. Sweeney, E. (1985). Clustering the *Science Citation Index* using cocitations I. A comparison of methods. *Scientometrics*, 7(3-6):391-409.

Smith, L. (1981). Citation analysis. *Library Trends*, 30(1):83-105.

Tijssen, R.J.W. (1992). *Cartography of science*: *scientometric mapping with multidimensional methods*. Leiden: DSWO Press, Leiden University.

Vladutz, G. & Cook, J. (1984). Bibliographic coupling and subject relatedness. *Proceedings of the ASIS Annual Meeting*, 47: 204-207.

Weinberg, B.H. (1974). Bibliographic coupling: a review. *Information Storage and Retrieval*. 10(5/&):189-195.

White, H. D. & Griffith, B. C. (1981). Author Cocitation: a literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32:163-171.

White, H. D. & McCain, K. W. (1998).Visualizing a Discipline: an author cocitation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*. 49(4): 327-335.

Ziman, John. (1984). An introduction to science studies: the philosophical and social aspects of science and technology. Cambridge: Cambridge University Press.

APPENDIX 1

EQUATIONS

In this appendix all equations are gathered and the context in which they occur in the text of the study is briefly quoted when motivated.

2.1

The number of r-combinations of a set of n distinct elements is denoted by $C(n, r)$ or $\binom{n}{r}$ and $C(n,r) = \frac{P(n,r)}{r!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{(n-r)!r!}$

This equation is applied to assess the share of all possible pairs of objects (articles or clusters) in a defined set that are bibliographically coupled (the number of coupled pairs divided by $C(n, r)$ where $r = 2$).[63] It was for instance applied to measure the density of matrixes of bibliographic couplings of the final document populations.

2.2

This equation is applied when assessing the degree of interconnectedness, the *density* (D), in sets of objects that may be described as *graphs*. D is defined as:

$$D = \frac{2 \cdot (\# L(G)}{N(N-1)},$$

where

$\#L(G)$= the number of edges connecting two vertices; and

N= the number of vertices (Otte & Rousseau, 2002).

The interval is [0, 1] and the maximum value is reached when the value of $\#L(G)$ equals the value of $N(N\text{-}1)/2$.

[63] This is similar to applying the next equation 2.2 though 2.2 is quoted and presented in a graphtheoretical context.

2.3

The Coupling Angle (C.A.) is expressed as:

$$C.A. = \frac{(D_{oj} \bullet D_{ok})}{\sqrt{(D_{oj} \cdot D_{oj})(D_{ok} \cdot D_{ok})}}$$

C.A. is the coupling angle for citing documents *j* and *k*. D_{oj} and D_{ok} are the binary vectors of document *j* respectively *k*. The C.A. takes the maximum value of 1 if two Boolean vectors are parallel and 0 if they are rectangular.

2.4

The *Jaccard coefficient* (commonly referred to as the Jaccard's index) is a well-known measure of the similarity *S* between two objects A and B, which counts the number of common attributes divided by the number of attributes possessed by at least one of the two objects:

$$S_{A,B} = \frac{|A \cap B|}{|A \cup B|}$$

In the context of cocitation analysis, this function is expressed as:

2.5

$$NCS_{ij} = \frac{C_{ij}}{(C_i + C_j - C_{ij})}$$

2.6

In the context of cocitation analysis, the *cosine function* (commonly referred to as Salton's cosine formula) is expressed as:

$$NCS_{ij} = \frac{C_{ij}}{\sqrt{(C_i \cdot C_j)}},$$

where:

NCS_{ij} = the normalized coupling strength between document *i* and *j*;

C_{ij} = the number of cocitations of document *i* and *j*;

C_i = the number of citations of document *i*; and

C_j = the number of citations of document *j*.

All three equations, 2.4-2.6, take values in the interval [0,1].

4.1

The C.A. (2.3) was in practice calculated as:

$$NCS_{ij} = \frac{r_{ij}}{(n_i \cdot n_j)^{1/2}},$$

where

NCS_{ij} = the normalized coupling strength between article i and article j

r_{ij} = number of references common to both i and j

n_i = number of references in the reference list of article i

n_j = number of references in the reference list of article j

The interval is [0, 1] and $n_i = n_j = r_{ij}$ gives the maximum value.

This equation is referred to as the normalized coupling strength (NCS) in the text.

4.2

A measure of the internal cluster coherence is the *Average Coupling Strength, AvgCS(C)*, for a cluster C. It is defined as:

$$AvgCS(C) = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} CS(d_i d_j)}{\binom{n}{2}},$$

where

n = number of articles in a cluster c,

CS = number of bibliographic coupling units between two articles, d_i, d_j

and

$d_i d_j (\in C)$

This equation is complementary to equation 2.2 as these two measures of cluster coherence reflect different aspects of internal cluster coherence.

4.3

This equation is applied for the measuring of the distance (similarity) between two clusters. When calculating the average distance between clusters in a set of clusters (resulting from a partition) changes of cluster isolation can be monitored.

Let C and C' be clusters of sizes k and m, respectively. *The average coupling strength between two clusters, C and C', AvgCS(C, C')*, is defined as:

$$AvgCS(C,C') = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} CS(d_i,d_j)}{k \times m},$$

where

CS = number of bibliographic coupling units between two articles, d_i, d_j

and $d_i \in C, d_j \in C'$

4.4

The concentration of articles to clusters was assessed applying *Pratt's measure of concentration.* This measure is of general use when one wants to see how concentrated or spread out items (here articles) are when partitioned into categories (here clusters).

Pratt's measure is given as:

$$C = \frac{2[((n+1)/2) - q]}{n-1},$$

where

C = Pratt's measure of concentration

n = number of categories

q = is the sum of rank times frequency for each category, divided by the total number of articles.

This measure will range between 0 and 1, where the most concentrated case (only one category) takes on the value of 1 and the "even" distribution the value of 0.

APPENDIX 2

BIBLIOGRAPHIC DESCRIPTIONS OF CLUSTERS WITH A SIZE ≥ 3 IN CASE 1

Bibliographic data of articles is presented in the following order: record number/ first author name/ publication year/ Journal name/ title/ author key words/key words plus. Missing data is indicated by "No Field".

CLUSTER 1

41/BURRELL QL/2002/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/ WILL THIS PAPER EVER BE CITED/NO FIELD/KEYWORDS PLUS: LIBRARY CIRCULATION MODEL

117/BURRELL QL/2002/SCIENTOMETRICS /THE NTH-CITATION DISTRIBUTION AND OBSOLESCENCE/NO FIELD/KEYWORDS PLUS: LIBRARY CIRCULATION MODEL

169/BURRELL QL/2001/ SCIENTOMETRICS / STOCHASTIC MODELLING OF THE FIRST-CITATION DISTRIBUTION/NO FIELD/KEYWORDS PLUS: LIBRARY CIRCULATION MODEL; OBSOLESCENCE; GROWTH

CLUSTER 3

36/CHEN CM/2002/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/VISUALIZING AND TRACKING THE GROWTH OF COMPETING PARADIGMS - 2 CASE-STUDIES/NO FIELD/KEYWORDS PLUS: AUTHOR COCITATION; INTELLECTUAL STRUCTURE; CO- CITATION; SCIENCE; SPACES; VIBE

176/SMALL H/2001/ SCIENTOMETRICS/BELVER AND HENRY/NO FIELD/KEYWORDS PLUS: SCIENTIFIC LITERATURES; CO-CITATION; SCIENCE

196/KOEHLER W/2001/ SCIENTOMETRICS/INFORMATION-SCIENCE AS LITTLE SCIENCE - THE IMPLICATIONS OF A BIBLIOMETRIC ANALYSIS OF THE JOURNAL-OF-THE-AMERICAN- SOCIETY-FOR-INFORMATION-SCIENCE/NO FIELD/KEYWORDS PLUS: SCIENTIFIC LITERATURE; CITATION ANALYSIS; AUTHORSHIP; LIBRARY; JASIS; COCITATION; COUNTRIES

210/JARNEVING B/2001/ SCIENTOMETRICS/THE COGNITIVE STRUCTURE OF CURRENT CARDIOVASCULAR RESEARCH/NO FIELD/KEYWORDS PLUS: SCIENTIFIC LITERATURES; SCIENCE

CLUSTER 4

19/CRONIN B/2001/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/HYPER AUTHORSHIP - A POST MODERN PERVERSION OR EVIDENCE OF A STRUCTURAL SHIFT IN SCHOLARLY COMMUNICATION PRACTICES/NO FIELD/KEYWORDS PLUS: MULTIPLE AUTHORSHIP; COLLABORATION; ARTICLES; SCIENCE; ORDER; ACKNOWLEDGMENTS; BIBLIOMETRICS; DISSEMINATION; CO-AUTHORSHIP; CONTRIBUTORS

45/CRONIN B/2001/ JOURNAL OF DOCUMENTATION/ACKNOWLEDGMENT TRENDS IN THE RESEARCH LITERATURE OF INFORMATION-SCIENCE/NO FIELD/KEYWORDS PLUS: COLLABORATION; SOCIOLOGY

106/CRONIN B/2002/ SCIENTOMETRICS/IDENTITY-CREATORS AND IMAGE-MAKERS - USING CITATION ANALYSIS AND THICK DESCRIPTION TO PUT AUTHORS IN THEIR PLACE/NO FIELD/KEYWORDS PLUS: SOCIOLOGY

CLUSTER 14

72/GARG KC/2002/ SCIENTOMETRICS/SCIENTOMETRICS OF LASER RESEARCH IN INDIA DURING 1970-1994/NO FIELD/KEYWORDS PLUS: SCIENCE; TECHNOLOGY; COLLABORATION; INDICATORS

80/GARG KC/2002/ SCIENTOMETRICS/SCIENTOMETRICS OF LASER RESEARCH IN INDIA AND CHINA/NO FIELD/KEYWORDS PLUS: CITATION PATTERNS; PUBLICATION; SCIENCE; IMPACT

110/HARITASH N/2002/ SCIENTOMETRICS/MAPPING OF S-AND-T ISSUES IN THE INDIAN PARLIAMENT - A SCIENTOMETRIC ANALYSIS OF QUESTIONS RAISED IN BOTH HOUSES OF THE PARLIAMENT/NO FIELD/KEYWORDS PLUS: INDICATORS; OUTPUT

189/GARG KC/2001/ SCIENTOMETRICS/A STUDY OF COLLABORATION IN LASER SCIENCE AND TECHNOLOGY/NO FIELD/KEYWORDS PLUS: INTERNATIONAL SCIENTIFIC COLLABORATION; POPULATION-GENETICS SPECIALITY; SCIENTOMETRICS

CLUSTER 16

76/LIANG LM/2002/ SCIENTOMETRICS/MAJOR FACTORS AFFECTING CHINA INTERREGIONAL RESEARCH COLLABORATION - REGIONAL SCIENTIFIC PRODUCTIVITY AND GEOGRAPHICAL PROXIMITY/NO FIELD/KEYWORDS PLUS: COOPERATION

163/STEFANIAK B/2001/ SCIENTOMETRICS/INTERNATIONAL-COOPERATION IN SCIENCE AND IN SOCIAL-SCIENCES AS REFLECTED IN MULTINATIONAL PAPERS INDEXED IN SCI AND SSCI/NO FIELD/KEYWORDS PLUS: COLLABORATION; COOPERATION; COUNTRIES; PROFILES

195/GLANZEL W/2001/ SCIENTOMETRICS/NATIONAL CHARACTERISTICS IN INTERNATIONAL SCIENTIFIC CO- AUTHORSHIP RELATIONS/NO FIELD/KEYWORDS PLUS: COLLABORATION

219/GLANZEL W/2001/ SCIENTOMETRICS/DOUBLE EFFORT = DOUBLE IMPACT - A CRITICAL-VIEW AT INTERNATIONAL CO-AUTHORSHIP IN CHEMISTRY/NO FIELD/KEYWORDS PLUS: SCIENTIFIC COLLABORATION; SCIENCES; MODEL

CLUSTER 18

4/LANGE LL/2002/ JOURNAL OF DOCUMENTATION/THE IMPACT FACTOR AS A PHANTOM - IS THERE A SELF-FULFILLING PROPHECY EFFECT OF IMPACT/AUTHOR KEYWORDS: VALUE ANALYSIS; ELECTRONIC PUBLISHING; DATABASES/KEYWORDS PLUS: JOURNAL IMPACT

131/VANLEEUWEN TN/2002/ SCIENTOMETRICS/DEVELOPMENT AND APPLICATION OF JOURNAL IMPACT MEASURES IN THE DUTCH SCIENCE SYSTEM/NO FIELD/KEYWORDS PLUS: CITATIONS; INSTITUTE

126/GLANZEL W/2002/ SCIENTOMETRICS/JOURNAL IMPACT MEASURES IN BIBLIOMETRIC RESEARCH/NO FIELD/KEYWORDS PLUS: SCIENTIFIC LITERATURE; STOCHASTIC-MODEL; CITATION; INDICATORS; PRODUCTIVITY; INDEX

CLUSTER 23

157/GURJEVA LG/2001/ SCIENTOMETRICS/SCIENTOMETRICS IN THE CONTEXT OF PROBABILISTIC PHILOSOPHY/NO FIELD/NO FIELD

160/NALIMOV VV/2001/ SCIENTOMETRICS/CITATION-CLASSICS OF NALIMOV,V.V - 1 - CURRENT-CONTENTS, NUMBER 21, MAY 21, 1990/NO FIELD/NO FIELD

161/NALIMOV VV/2001/ SCIENTOMETRICS/CITATION-CLASSICS OF NALIMOV,V.V. 2 - CURRENT-CONTENTS, NUMBER 24, JUNE 11, 1990

168/SHAPIRO SI/2001/ SCIENTOMETRICS/THE UNIVERSE GRASPER/NO FIELD/NO FIELD

CLUSTER 27

11/HUBER JC/2001/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/A NEW METHOD FOR ANALYZING SCIENTIFIC PRODUCTIVITY/NO FIELD/KEYWORDS PLUS: STATIONARY SCIENTOMETRIC DISTRIBUTIONS; CUMULATIVE ADVANTAGE; CREATIVITY; PARTICIPATION; PUBLICATION; STATISTICS; DURATION; SPEED

40/HUBER JC/2002/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/A NEW MODEL THAT GENERATES LOTKAS LAW/NO FIELD/KEYWORDS PLUS: SUCCESS-BREEDS-SUCCESS; INFORMETRIC DISTRIBUTIONS; CUMULATIVE ADVANTAGE; SCIENTIFIC PRODUCTIVITY; INVENTIVE PRODUCTIVITY; STATISTICS; RANDOMNESS; CREATIVITY; PUBLICATION; EXCEEDANCES

214/HUBER JC/2001/ SCIENTOMETRICS/SCIENTIFIC PRODUCTION - A STATISTICAL-ANALYSIS OF AUTHORS IN PHYSICS, 1800-1900/NO FIELD/NO FIELD

227/HUBER JC/2001/ SCIENTOMETRICS/SCIENTIFIC PRODUCTION - A STATISTICAL-ANALYSIS OF AUTHORS IN MATHEMATICAL LOGIC/NO FIELD/KEYWORDS PLUS: STATIONARY SCIENTOMETRIC DISTRIBUTIONS; CUMULATIVE ADVANTAGE; PARTICIPATION; PUBLICATION; DURATION; TESTS; SPEED; LAW

CLUSTER 28

12/IVANCHEVA LE/2001/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/THE NON-GAUSSIAN NATURE OF BIBLIOMETRIC AND SCIENTOMETRIC DISTRIBUTIONS - A NEW APPROACH TO INTERPRETATION/NO FIELD/KEYWORDS PLUS: PRODUCTIVITY

18/KRETSCHMER H/2001/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/AUTHOR INFLATION LEADS TO A BREAKDOWN OF LOTKAS LAW/NO FIELD/KEYWORDS PLUS: INFORMETRIC DISTRIBUTIONS; SCIENTIFIC COLLABORATION; PRODUCTIVITY; ATTRIBUTION; COUNTS

137/KARISIDDAPPA CR/2002/ SCIENTOMETRICS/SCIENTIFIC PRODUCTIVITY OF AUTHORS IN THEORETICAL POPULATION- GENETICS/NO FIELD/KEYWORDS PLUS: FREQUENCY-DISTRIBUTION; LOTKAS LAW; PUBLICATION; TIME

CLUSTER 36

94/LARSEN B/2002/ SCIENTOMETRICS/EXPLOITING CITATION OVERLAPS FOR INFORMATION-RETRIEVAL - GENERATING A BOOMERANG EFFECT FROM THE NETWORK OF SCIENTIFIC PAPERS/NO FIELD/KEYWORDS PLUS: SYSTEMS; SCIENCE; DESIGN; WEB

183/SANDSTROM PE/2001/ SCIENTOMETRICS/SCHOLARLY COMMUNICATION AS A SOCIOECOLOGICAL SYSTEM/NO FIELD/KEYWORDS PLUS: HUMAN BEHAVIORAL ECOLOGY; CO-CITATION; SCIENTIFIC LITERATURES; INTELLECTUAL STRUCTURE; INFORMATION- SEEKING; AUTHOR COCITATION; SCIENCE; RETRIEVAL; DOCUMENTS; SPACE

184/WHITE HD/2001/ SCIENTOMETRICS/AUTHOR-CENTERED BIBLIOMETRICS THROUGH CAMEOS - CHARACTERIZATIONS AUTOMATICALLY MADE AND EDITED ONLINE/NO FIELD/KEYWORDS PLUS: CITATION ANALYSIS; PUBLICATIONS; RETRIEVAL; MODEL

CLUSTER 38

10/LEYDESDORFF L/2001/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/THE SELF-ORGANIZATION OF THE EUROPEAN INFORMATION-SOCIETY - THE CASE OF BIOTECHNOLOGY/NO FIELD/KEYWORDS PLUS: CO-CITATIONS; SCIENCE; INDICATORS; GOVERNMENT; TECHNOLOGY; INDUSTRY; WORDS

30/LEYDESDORFF L/2002/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/DYNAMIC AND EVOLUTIONARY UPDATES OF CLASSIFICATORY SCHEMES IN SCIENTIFIC JOURNAL STRUCTURES/NO FIELD/KEYWORDS PLUS: BRITISH SCIENCE; BIBLIOMETRIC ASSESSMENT; DECLINE; PERFORMANCE; INDICATORS; NATIONS

141/LEYDESDORFF L/2002/ SCIENTOMETRICS/INDICATORS OF STRUCTURAL-CHANGE IN THE DYNAMICS OF SCIENCE - ENTROPY STATISTICS OF THE SCI JOURNAL-CITATION-REPORTS/NO FIELD/KEYWORDS PLUS: COMMUNICATION; INTELLIGENCE; PERFORMANCE; TECHNOLOGY; KNOWLEDGE; IMPACT; AREAS

220/VILANOVA MR/2001/ SCIENTOMETRICS/WHY CATALONIA CANNOT BE CONSIDERED AS A REGIONAL INNOVATION SYSTEM/NO FIELD/KEYWORDS PLUS: INDUSTRY-GOVERNMENT RELATIONS; PATENT STATISTICS; EUROPEAN-UNION; TRIPLE-HELIX; SCIENCE; TECHNOLOGY

CLUSTER 41

89/VERBEEK A/2002/ SCIENTOMETRICS/LINKING SCIENCE TO TECHNOLOGY - USING BIBLIOGRAPHIC REFERENCES IN PATENTS TO BUILD LINKAGE SCHEMES/NO FIELD/KEYWORDS PLUS: INDICATORS

96/MEYER M/2002/ SCIENTOMETRICS/TRACING KNOWLEDGE FLOWS IN INNOVATION SYSTEMS/NO FIELD/KEYWORDS PLUS: PATENT CITATIONS; SCIENCE; TECHNOLOGY; INDICATORS; INVENTIONS; INDUSTRY; LINKAGE; US

199/MEYER MS/2001/ SCIENTOMETRICS/PATENT CITATION ANALYSIS IN A NOVEL FIELD OF TECHNOLOGY - AN EXPLORATION OF NANO-SCIENCE AND NANO-TECHNOLOGY/NO FIELD/KEYWORDS PLUS: TECHNICAL CHANGE

CLUSTER 45

101/PERITZ BC/2002/ SCIENTOMETRICS/THE SOURCES USED BY BIBLIOMETRICS-SCIENTOMETRICS AS REFLECTED IN REFERENCES/NO FIELD/KEYWORDS PLUS: SCIENTIFIC JOURNALS; CITATION ANALYSIS; SELF- CITATION; SCIENCE; PATTERNS; DECADES

133/SCHUBERT A/2002/ SCIENTOMETRICS/THE WEB OF SCIENTOMETRICS - A STATISTICAL OVERVIEW OF THE 1ST 50 VOLUMES OF THE JOURNAL/NO FIELD/NO FIELD

225/SCHOEPFLIN U/2001/ SCIENTOMETRICS/2 DECADES OF SCIENTOMETRICS - AN INTERDISCIPLINARY FIELD REPRESENTED BY ITS LEADING JOURNAL/NO FIELD/KEYWORDS PLUS: SCIENCES

CLUSTER 47.

23/WHITE HD/2001/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/AUTHORS AS CITERS OVER TIME/NO FIELD/KEYWORDS PLUS: CITATION ANALYSIS; ORTEGA HYPOTHESIS; INFORMATION-SCIENCE; SELF-CITATIONS; MOTIVATIONS; KNOWLEDGE; MODEL; CLASSIFICATION; DOCUMENTATION; REFERENCES

25/WHITLEY KM/2002/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/ANALYSIS OF SCIFINDER SCHOLAR AND WEB OF SCIENCE CITATION SEARCHES/NO FIELD/NO FIELD

102/PICHAPPAN P/2002/ SCIENTOMETRICS/THE OTHER SIDE OF THE COIN - THE INTRICACIES OF AUTHOR SELF- CITATIONS/NO FIELD/KEYWORDS PLUS: SCIENCE; COMMUNICATION; BEHAVIOR; LEVEL

CLUSTER 48

103/PRIME C/2002/ SCIENTOMETRICS/COCITATIONS AND CO-SITATIONS - A CAUTIONARY VIEW ON AN ANALOGY/NO FIELD/KEYWORDS PLUS: SCIENTIFIC LITERATURE; SCIENCE; COCITATION; IMPACT

206/SCHWECHHEIMER H/2001/ SCIENTOMETRICS/MAPPING INTERDISCIPLINARY RESEARCH FRONTS IN NEUROSCIENCE - A BIBLIOMETRIC VIEW TO RETROGRADE-AMNESIA/NO FIELD/KEYWORDS PLUS: CO-CITATIONS; SCIENCE

224/SALZARULO L/2001/ SCIENTOMETRICS/BIAS, STRUCTURE AND QUALITY IN CITATION INDEXING/NO FIELD/KEYWORDS PLUS: CO-CITATIONS; SCIENCE

CLUSTER 51

2/THELWALL M/2002/JOURNAL OF DOCUMENTATION/EVIDENCE FOR THE EXISTENCE OF GEOGRAPHIC TRENDS IN UNIVERSITY WEB SITE INTERLINKING/AUTHOR KEYWORDS: INTERNET; KNOWLEDGE WORKERS; UNIVERSITIES; UNITED KINGDOM/KEYWORDS PLUS: CITATION ANALYSIS; IMPACT FACTORS; INFORMATION; INTERNET; SCIENCE; CRAWLER

5/THELWALL M/2002/ JOURNAL OF DOCUMENTATION /A COMPARISON OF SOURCES OF LINKS FOR ACADEMIC WEB IMPACT FACTOR CALCULATIONS/AUTHOR KEYWORDS: INTERNET; INFORMATION RETRIEVAL/KEYWORDS PLUS: CITATION; INFORMATION

8/THELWALL M/2001/JOURNAL OF INFORMATION SCIENCE/EXPLORING THE LINK STRUCTURE OF THE WEB WITH NETWORK DIAGRAMS/NO FIELD/KEYWORDS PLUS: IMPACT FACTORS; SEARCH

15/THELWALL M/2001/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/EXTRACTING MACROSCOPIC INFORMATION FROM WEB LINKS/NO FIELD/KEYWORDS PLUS: RESEARCH ASSESSMENT EXERCISE; WORLD-WIDE-WEB; IMPACT FACTORS; SCHOLARLY COMMUNICATION; UNIVERSITY DEPARTMENTS; CITATION COUNTS; SEARCH ENGINE; BRITISH; CONTINUUM; ANATOMY

31/THELWALL M/2002/ JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY/CONCEPTUALIZING DOCUMENTATION ON THE WEB - AN EVALUATION OF DIFFERENT HEURISTIC-BASED MODELS FOR COUNTING LINKS BETWEEN UNIVERSITY WEB SITES/NO FIELD/KEYWORDS PLUS: IMPACT FACTORS; CITATION ANALYSIS; SEARCH ENGINE; COMMUNICATION; INTERNET; INFORMATION; CRAWLER; DESIGN; PAGES

46/THELWALL M/2001/ JOURNAL OF INFORMATION SCIENCE /A WEB CRAWLER DESIGN FOR DATA MINING/NO FIELD/KEYWORDS PLUS: IMPACT FACTORS; SEARCH ENGINE; SITE

62/THELWALL M/2002/ SCIENTOMETRICS/INTERLINKING BETWEEN ASIA-PACIFIC UNIVERSITY WEB SITES/NO FIELD/NO FIELD

87/SMITH A/2002/ SCIENTOMETRICS/WEB IMPACT FACTORS FOR AUSTRALASIAN UNIVERSITIES/NO FIELD/KEYWORDS PLUS: CO-AUTHORSHIP

CLUSTER 62

144/BEAVER DD/2001/ SCIENTOMETRICS/REFLECTIONS ON SCIENTIFIC COLLABORATION, (AND ITS STUDY) - PAST, PRESENT, AND FUTURE/NO FIELD/KEYWORDS PLUS: CO-AUTHORSHIP

63/MARTINSEMPERE MJ/2002/ SCIENTOMETRICS /THE EFFECT OF TEAM CONSOLIDATION ON RESEARCH COLLABORATION AND PERFORMANCE OF SCIENTISTS - CASE-STUDY OF SPANISH UNIVERSITY RESEARCHERS IN GEOLOGY/NO FIELD/NO FIELD

69/FARAHAT H/2002/ SCIENTOMETRICS /AUTHORSHIP PATTERNS IN AGRICULTURAL SCIENCES IN EGYPT/NO FIELD/KEYWORDS PLUS: SCIENTIFIC CO-AUTHORSHIP; RESEARCH COLLABORATION; MULTIPLE AUTHORSHIP; LIBRARY

153/WAGNERDOBLER R/2001/ SCIENTOMETRICS /CONTINUITY AND DISCONTINUITY OF COLLABORATION BEHAVIOR SINCE 1800 - FROM A BIBLIOMETRIC POINT-OF-VIEW/NO FIELD/KEYWORDS PLUS: SCIENTIFIC CO-AUTHORSHIP

APPENDIX 3

THE COMPARISON OF TWO PARTITIONS IN CASE 1

The Dispersion of Articles over Clusters for Two Partitions.

In the following table, columns A-D show the dispersion of articles in clusters generated by the field expert over the clusters generated by the complete link cluster method whereas columns E-H show the dispersion of articles in clusters generated by the complete link cluster method over the clusters generated by the field expert.

A	B	C	D	E	F	G	H
Complete doc.nr.	Complete clu.nr.	Expert doc.nr.	Expert clu.nr.	Complete doc.nr.	Complete clu.nr.	Expert doc.nr.	Expert clu.nr.
4	18	4	1	41	1	41	2
126	18	126	1	117	1	117	2
131	18	131	1	169	1	169	2
30	38	30	1	36	3	36	3
141	38	141	1	210	3	210	3
25	47	25	1	196	3	196	8
224	48	224	1	176	3	176	9
41	1	41	2	19	4	19	4
117	1	117	2	106	4	106	7
169	1	169	2	45	4	45	8
11	27	11	2	110	14	110	6
40	27	40	2	72	14	72	8
214	27	214	2	80	14	80	8
227	27	227	2	189	14	189	8
12	28	12	2	76	16	76	4
18	28	18	2	163	16	163	4
36	3	36	3	195	16	195	4
210	3	210	3	219	16	219	4
183	36	183	3	4	18	4	1
184	36	184	3	126	18	126	1
206	48	206	3	131	18	131	1
19	4	19	4	157	23	157	9
76	16	76	4	160	23	160	9
163	16	163	4	161	23	161	9
195	16	195	4	168	23	168	9
219	16	219	4	11	27	11	2
63	62	63	4	40	27	40	2
69	62	69	4	214	27	214	2
144	62	144	4	227	27	227	2
153	62	153	4	12	28	12	2
133	45	133	5	18	28	18	2
103	48	103	5	137	28	137	8
2	51	2	5	94	36	94	10
5	51	5	5	183	36	183	3
8	51	8	5	184	36	184	3
15	51	15	5	30	38	30	1
31	51	31	5	141	38	141	1
46	51	46	5	10	38	10	6

62	51	62	5
87	51	87	5
110	14	110	6
10	38	10	6
220	38	220	6
89	41	89	6
96	41	96	6
199	41	199	6
106	4	106	7
23	47	23	7
102	47	102	7
196	3	196	8
45	4	45	8
72	14	72	8
80	14	80	8
189	14	189	8
137	28	137	8
101	45	101	8
225	45	225	8
176	3	176	9
157	23	157	9
160	23	160	9
161	23	161	9
168	23	168	9
94	36	94	10
220	38	220	6
89	41	89	6
96	41	96	6
199	41	199	6
133	45	133	5
101	45	101	8
225	45	225	8
25	47	25	1
23	47	23	7
102	47	102	7
224	48	224	1
206	48	206	3
103	48	103	5
2	51	2	5
5	51	5	5
8	51	8	5
15	51	15	5
31	51	31	5
46	51	46	5
62	51	62	5
87	51	87	5
63	62	63	4
69	62	69	4
144	62	144	4
153	62	153	4

APPENDIX 4:

BIBLIOGRAPHIC DESCRIPTIONS OF CLUSTERS WITH A SIZE ≥ 3 IN CASE 2

Bibliographic data as follows: record number/ first author name/ publication year/ Journal name/ title/ author key words/key words plus. Missing data is indicated by "No Field".

CLUSTER 1

1014/SHI M/2002/ ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/CATALYTIC, ASYMMETRIC BAYLIS-HILLMAN REACTION OF IMINES WITH METHYL VINYL KETONE AND METHYL ACRYLATE/NO FIELD/KEYWORDS PLUS: TITANIUM(IV) CHLORIDE; LEWIS BASE; ALDEHYDES; PHOSPHINE; ESTERS

7888/SHI M/2003/ JOURNAL OF ORGANIC CHEMISTRY/AN UNEXPECTED HIGHLY STEREOSELECTIVE DOUBLE AZA-BAYLIS- HILLMAN REACTION OF SULFONATED IMINES WITH PHENYL VINYL KETONE/NO FIELD/KEYWORDS PLUS: LEWIS BASE; TITANIUM(IV) CHLORIDE; ALDEHYDES; PHOSPHINE

12254/SHI M/2002/ TETRAHEDRON LETTERS/ONE-POT AZA-BAYLIS-HILLMAN REACTIONS OF ARYLALDEHYDES AND DIPHENYLPHOSPHINAMIDE WITH METHYL VINYL KETONE IN THE PRESENCE OF TICL4PPH3, AND ET3N/NO FIELD/KEYWORDS PLUS: ELECTRON-DEFICIENT ALKENES; TITANIUM(IV) CHLORIDE; LEWIS BASE; ALDEHYDES; PHOSPHINE; ESTERS

13307/SHI M/2002/ TETRAHEDRON LETTERS/BAYLIS-HILLMAN REACTIONS OF N-ARYLIDENEDIPHENYLPHOSPHINAMIDES WITH METHYL VINYL KETONE, METHYL ACRYLATE, AND ACRYLONITRILE/AUTHOR KEYWORDS: N-ARYLIDENEDIPHENYLPHOSPHINAMIDE; LEWIS BASE; BAYLIS-HILLMAN REACTION; METHYL VINYL KETONE (MVK); METHYL ACRYLATE; ACRYLONITRILE/KEYWORDS PLUS: ELECTRON-DEFICIENT ALKENES; TITANIUM(IV) CHLORIDE; LEWIS BASE; ALDEHYDES; PHOSPHINE; ESTERS

14376/SHI M/2002/ TETRAHEDRON LETTERS/LEWIS BASE AND L-PROLINE CO-CATALYZED BAYLIS-HILLMAN REACTION OF ARYLALDEHYDES WITH METHYL VINYL KETONE/AUTHOR KEYWORDS: BAYLIS-HILLMAN REACTION; LEXIS BASE; METHYL VINYL KETONE (MVK); L-PROLINE; IMIDAZOLE; TRIETHYLAMINE/KEYWORDS PLUS: TITANIUM(IV) CHLORIDE; CONJUGATE ADDITION; ALDOL REACTIONS; ALDEHYDES

CLUSTER 2

7415/CASTRO EA/2003/ JOURNAL OF ORGANIC CHEMISTRY/KINETIC INVESTIGATION OF THE REACTIONS OF S-4-NITROPHENYL 4- SUBSTITUTED THIOBENZOATES WITH SECONDARY ALICYCLIC AMINES IN AQUEOUS-ETHANOL/NO FIELD/KEYWORDS PLUS: NUCLEOPHILIC-SUBSTITUTION REACTIONS; S-ARYL THIOCARBONATES; 4-NITROPHENYL THIONOCARBONATES; METHYL CARBONATE; ESTER AMINOLYSIS; MECHANISM; ACETONITRILE; 2,4- DINITROPHENYL; PHENYL; PYRIDINOLYSIS

7732/CASTRO EA/2003/ JOURNAL OF ORGANIC CHEMISTRY/KINETICS AND MECHANISM OF THE AMINOLYSIS OF 4-METHYLPHENYL AND 4-CHLOROPHENYL 2,4-DINITROPHENYL CARBONATES IN AQUEOUS- ETHANOL/NO FIELD/KEYWORDS PLUS: STRUCTURE-REACTIVITY CORRELATIONS; 2,4,6- TRINITROPHENYL METHYL CARBONATE; SECONDARY ALICYCLIC AMINES; S-ARYL THIOLCARBONATES; CONCERTED MECHANISM; SUBSTITUTED PYRIDINES; TRANSITION-STATE; ESTER AMINOLYSIS; ETHYL; ACETATE

8686/CASTRO EA/2002/ JOURNAL OF ORGANIC CHEMISTRY/KINETICS AND MECHANISM OF THE AMINOLYSIS OF METHYL 4- NITROPHENYL, METHYL 2,4-DINITROPHENYL, AND PHENYL 2,4-DINITROPHENYL CARBONATES/NO FIELD/KEYWORDS PLUS: SECONDARY ALICYCLIC AMINES; STRUCTURE- REACTIVITY CORRELATIONS; RATE-DETERMINING STEP; 2,4,6- TRINITROPHENYL ACETATE; SUBSTITUTED PYRIDINES; CONCERTED MECHANISMS; AQUEOUS-SOLUTION; PYRIDINOLYSIS; THIONOCARBONATES; THIOCARBONATE

CLUSTER 3

7672/CASTRO EA/2003/JOURNAL OF ORGANIC CHEMISTRY/KINETIC-STUDY OF THE PHENOLYSIS OF O-METHYL AND O-PHENYL O- 2,4-DINITROPHENYL THIOCARBONATES AND O-ETHYL 2,4-DINITROPHENYL DITHIOCARBONATE/NO FIELD/KEYWORDS PLUS: STRUCTURE-REACTIVITY CORRELATIONS; PHENOLATE ION NUCLEOPHILES; SUBSTITUTED PHENOXIDE IONS; SECONDARY ALICYCLIC AMINES; TRANSITION-STATE STRUCTURE; ACYL-TRANSFER REACTIONS; ACETYL GROUP TRANSFER; CONCERTED MECHANISMS; 4- NITROPHENYL CHLOROTHIONOFORMATES; OXYGEN NUCLEOPHILES

8098/CASTRO EA/2003/JOURNAL OF ORGANIC CHEMISTRY/KINETICS AND MECHANISM OF THE BENZENETHIOLYSIS OF 2,4- DINITROPHENYL AND 2,4,6-TRINITROPHENYL METHYL CARBONATES AND S-(2,4-DINITROPHENYL) AND S-(2,4,6-TRINITROPHENYL) ETHYL THIOLCARBONATES/NO FIELD/KEYWORDS PLUS: ACYL GROUP TRANSFER; STRUCTURE-REACTIVITY CORRELATIONS; SUBSTITUTED PHENOXIDE IONS; S-ARYL THIOCARBONATES; CONCERTED MECHANISMS; TRANSITION-STATE; AQUEOUS-SOLUTION; OXYGEN NUCLEOPHILES; MECN-H2O MIXTURES; ESTER AMINOLYSIS

9348/CASTRO EA/2002/JOURNAL OF ORGANIC CHEMISTRY/KINETICS AND MECHANISM OF THE PHENOLYSIS OF ASYMMETRIC DIARYL CARBONATES/NO FIELD/KEYWORDS PLUS: STRUCTURE-REACTIVITY CORRELATIONS; PHENOLATE ION NUCLEOPHILES; SECONDARY ALICYCLIC AMINES; SUBSTITUTED PHENOXIDE IONS; TRANSITION-STATE STRUCTURE; ACYL-TRANSFER REACTIONS; ACETYL GROUP TRANSFER; CONCERTED MECHANISMS; AQUEOUS-SOLUTION; AMINOLYSIS

CLUSTER 4

10297/LHOTAK P/2003/TETRAHEDRON LETTERS/SYNTHESIS OF A DEEP-CAVITY THIACALIX(4)ARENE/AUTHOR KEYWORDS: THIACALIXARENES; X-RAY CRYSTALLOGRAPHY; ALKYLATION; CONFORMATIONAL ANALYSIS/KEYWORDS PLUS: STATE STRUCTURAL-ANALYSIS; SOLID-STATE; INFINITE CHANNELS; CALIX(4)ARENES; DERIVATIVES; RIM; P-TERT-BUTYLTHIACALIX(4)ARENE; THIACALIXARENE; CALIXARENES; CONFORMERS

10480/LHOTAK P/2003/TETRAHEDRON LETTERS/STEREOSELECTIVE OXIDATION OF THIACALIX(4)ARENES WITH THE NANO3/CF3COOH SYSTEM/NO FIELD/KEYWORDS PLUS: SELECTIVE OXIDATION; METAL-IONS; UPPER RIM; P- TERT-BUTYLTHIACALIX(4)ARENE; SULFINYLCALIX(4)ARENES; CALIX(4)ARENES; DERIVATIVES; SULFINYL

12144/LHOTAK P/2002/TETRAHEDRON LETTERS/ALKYLATION OF THIACALIX(4)ARENES/NO FIELD/KEYWORDS PLUS: SOLID-STATE; INFINITE CHANNELS; P-TERT-BUTYLTHIACALIX(4)ARENE; CONFORMERS

12650/LHOTAK P/2002/TETRAHEDRON LETTERS/NITRATION OF THIACALIX(4)ARENE DERIVATIVES/NO FIELD/KEYWORDS PLUS: SELECTIVE OXIDATION; METAL-IONS; P-TERT-BUTYLTHIACALIX(4)ARENE; SULFINYLCALIX(4)ARENES; CALIXARENES

13490/LHOTAK P/2002/TETRAHEDRON LETTERS/DIAZO COUPLING - AN ALTERNATIVE METHOD FOR THE UPPER RIM AMINATION OF THIACALIX(4)ARENES/NO FIELD/KEYWORDS PLUS: P-TERT-BUTYLTHIACALIX(4)ARENE; CALIXARENES; OXIDATION

CLUSTER 5

6342/LEWIS FD/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/DYNAMICS OF INTERSTRAND AND INTRASTRAND HOLE TRANSPORT IN DNA HAIRPINS/NO FIELD/KEYWORDS PLUS: CHARGE-TRANSPORT; DISTANCE; MECHANISM; OXIDATION

11268/TAKADA T/2003/TETRAHEDRON LETTERS/HOLE TRANSFER IN DNA - DNA AS A SCAFFOLD FOR HOLE TRANSFER BETWEEN 2 ORGANIC-MOLECULES/NO FIELD/KEYWORDS PLUS: DISTANCE CHARGE-TRANSPORT; TRANSIENT ABSORPTION; HOPPING MECHANISM; ELECTRON-TRANSFER; REDOX CHEMISTRY; RADICAL-CATION; DUPLEX DNA; OXIDATION; DYNAMICS; 2-AMINOPURINE

12470/KAWAI K/2002/TETRAHEDRON LETTERS/REGULATION OF ONE-ELECTRON OXIDATION RATE OF GUANINE AND HOLE TRANSFER RATE IN DNA THROUGH HYDROGEN-BONDING/AUTHOR KEYWORDS: DNA; HYDROGEN BONDING; ONE-ELECTRON OXIDATION; HOLE TRANSFER/KEYWORDS PLUS: DIIMIDE DERIVATIVES; TRANSPORT; DISTANCE; SEQUENCE; NAPHTHALENE; CHEMISTRY; CLEAVAGE; IMIDE; GG

CLUSTER 6

723/GARTNER ZJ/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/2 ENABLING ARCHITECTURES FOR DNA-TEMPLATED ORGANIC-SYNTHESIS/AUTHOR KEYWORDS: BIOORGANIC CHEMISTRY; COMBINATORIAL CHEMISTRY; DNA; TEMPLATE SYNTHESIS/KEYWORDS PLUS: NUCLEIC-ACIDS; LIGATION; OLIGONUCLEOTIDES; AMPLIFICATION; REPLICATION; SYSTEM; ORIGIN; PNA

1112/CALDERONE CT/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/DIRECTING OTHERWISE INCOMPATIBLE REACTIONS IN A SINGLE SOLUTION BY USING DNA-TEMPLATED ORGANIC-SYNTHESIS/AUTHOR KEYWORDS: COMBINATORIAL CHEMISTRY; DIVERSIFICATION; OLIGONUCLEOTIDES; SYNTHETIC METHODS; TEMPLATE SYNTHESIS/KEYWORDS PLUS: NATURAL PRODUCT; LIGATION; AMPLIFICATION; MOLECULES

1579/GARTNER ZJ/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/EXPANDING THE REACTION SCOPE OF DNA-TEMPLATED SYNTHESIS/AUTHOR KEYWORDS: COUPLING REACTIONS; DNA; MOLECULAR EVOLUTION; SYNTHETIC METHODS; TEMPLATE SYNTHESIS/KEYWORDS PLUS: REPLICATION; LIGATION; ACIDS; RNA

CLUSTER 7

495/PIDATHALA C/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/DIRECT CATALYTIC ASYMMETRIC ENOLEXO ALDOLIZATIONS/AUTHOR KEYWORDS: ALDOL REACTION; AMINO ACIDS; ASYMMETRIC CATALYSIS; ORGANOCATALYSIS/KEYWORDS PLUS: 3-COMPONENT MANNICH REACTION; DYNAMIC KINETIC RESOLUTION; ALDOL REACTIONS; AMINO-ACIDS; CARBOXYLIC ESTERS; ALPHA-AMINATION; PROLINE; ALDEHYDES; INDUCTION; HYDROGENATION

4105/BAHMANYAR S/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/QUANTUM-MECHANICAL PREDICTIONS OF THE STEREOSELECTIVITIES OF PROLINE-CATALYZED ASYMMETRIC INTERMOLECULAR ALDOL REACTIONS/NO FIELD/KEYWORDS PLUS: MOLECULAR-ORBITAL METHODS; GAUSSIAN-TYPE BASIS; ORGANIC-MOLECULES

4502/HOANG L/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/KINETIC AND STEREOCHEMICAL EVIDENCE FOR THE INVOLVEMENT OF ONLY ONE PROLINE MOLECULE IN THE TRANSITION-STATES OF PROLINE-CATALYZED INTRAMOLECULAR AND INTERMOLECULAR ALDOL REACTIONS/NO FIELD/KEYWORDS PLUS: ASYMMETRIC-SYNTHESIS; MECHANISM; CYCLIZATION; KETONES

CLUSTER 8

12402/WALSH LM/2002/TETRAHEDRON LETTERS/SULFIDE-BF3-CENTER-DOT-OET2 MEDIATED BAYLIS-HILLMAN REACTIONS/NO FIELD/KEYWORDS PLUS: BETA-HYDROXY-KETONES; TITANIUM(IV) CHLORIDE; ALPHA,BETA-UNSATURATED KETONES; ALDEHYDES; CATALYSTS

12575/CATRI R/2002/TETRAHEDRON LETTERS/IMIDAZOLE-CATALYZED BAYLIS-HILLMAN REACTIONS - A NEW ROUTE TO ALLYLIC ALCOHOLS FROM ALDEHYDES AND CYCLIC ENONES/NO FIELD/NO FIELD

12730/KATAOKA T/2002/TETRAHEDRON LETTERS/TANDEM MICHAEL-ALDOL REACTION VIA 6-ENDO-DIG CYCLIZATION OF YNONE-CHALCOGENIDES - SYNTHESIS OF 2-UNSUBSTITUTED 3-(HYDROXYALKYL)CHALCOGENOCHROMEN-4-ONES/NO FIELD/KEYWORDS PLUS: BAYLIS-HILLMAN REACTION; ELECTRON-DEFICIENT ALKENES; ALPHA,BETA-ACETYLENIC KETONES; ETHYNYL KETONES; ALDEHYDES; FLAVONES

CLUSTER 9

534/STORER RI/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/A TOTAL-SYNTHESIS OF EPOTHILONES USING SOLID-SUPPORTED REAGENTS AND SCAVENGERS/AUTHOR KEYWORDS: ALDOL REACTION; ANTITUMOR AGENTS; NATURAL PRODUCTS; POLYKETIDES; POLYMERS/KEYWORDS PLUS: ASYMMETRIC ALDOL REACTION; ENANTIOSELECTIVE TOTAL SYNTHESIS; STEREOSELECTIVE TOTAL SYNTHESIS; MICROTUBULE-STABILIZING AGENTS; OLEFIN METATHESIS APPROACH; FORMAL TOTAL SYNTHESIS; CHIRAL LEWIS-ACID; ALKYNE METATHESIS; ORGANIC-SYNTHESIS; SOLUTION-PHASE

1631/SUN J/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/STEREOSELECTIVE TOTAL SYNTHESIS OF EPOTHILONES BY THE METATHESIS APPROACH INVOLVING C9-C10 BOND FORMATION/NO FIELD/KEYWORDS PLUS: MICROTUBULE-STABILIZING AGENTS; OLEFIN METATHESIS; B ANALOGS; TAXOL; CELLS

1640/LIU JJ/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/ALDOLASE-CATALYZED ASYMMETRIC-SYNTHESIS OF NOVEL PYRANOSE SYNTHONS AS A NEW ENTRY TO HETEROCYCLES AND EPOTHILONES/AUTHOR KEYWORDS: ALDOL REACTION; ENZYME CATALYSIS; EPOTHILONES; SYNTHETIC METHODS; TOTAL SYNTHESIS/KEYWORDS PLUS: ENANTIOSELECTIVE TOTAL SYNTHESIS; 2- DEOXYRIBOSE-5-PHOSPHATE ALDOLASE; STEREOSELECTIVE SYNTHESIS; CHEMOENZYMATIC SYNTHESIS; (-)-EPOTHILONE-A; CONFORMATION; KETONES

8856/CHAPPELL MD/2002/JOURNAL OF ORGANIC CHEMISTRY/PROBING THE SAR OF DEPOB VIA CHEMICAL SYNTHESIS - A TOTAL SYNTHESIS EVALUATION OF C26-(1,3-DIOXOLANYL)-12,13-DESOXYEPOTHILONE-B/NO FIELD/KEYWORDS PLUS: STEREOCONTROLLED TOTAL SYNTHESIS; ENANTIOSELECTIVE TOTAL SYNTHESIS; OLEFIN METATHESIS APPROACH; CROSS-COUPLING REACTIONS; DRUG DISCOVERY PROCESS; SIDE- CHAIN ANALOGS; EPOTHILONE-B; ASYMMETRIC DIHYDROXYLATION; BIOLOGICAL EVALUATION; 12,13-CYCLOBUTYL EPOTHILONES

13651/ERMOLENKO MS/2002/TETRAHEDRON LETTERS/SYNTHESIS OF EPOTHILONE-B AND EPOTHILONE-D FROM D-GLUCOSE/NO FIELD/KEYWORDS PLUS: ENANTIOSELECTIVE TOTAL SYNTHESIS; OLEFIN METATHESIS APPROACH; STEREOSELECTIVE SYNTHESIS; SORANGIUM-CELLULOSUM; ALDOL CONDENSATION; ANALOGS; (-)-EPOTHILONE-A; 12,13-DESOXYEPOTHILONE-B; DERIVATIVES; DISCOVERY

CLUSTER 10

7018/LIST B/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/THE PROLINE-CATALYZED DIRECT ASYMMETRIC 3-COMPONENT MANNICH REACTION - SCOPE, OPTIMIZATION, AND APPLICATION TO THE HIGHLY ENANTIOSELECTIVE SYNTHESIS OF 1,2-AMINO ALCOHOLS/NO FIELD/KEYWORDS PLUS: LITHIUM ESTER ENOLATE; GLYCOL ALDEHYDE HYDRAZONES; CHIRAL ZIRCONIUM CATALYST; VICINAL AMINO- ALCOHOLS; ALPHA-IMINO ESTERS; BETA-AMINO; ALLIBIS(BINAPHTHOXIDE) COMPLEX; DIASTEREOSELECTIVE SYNTHESIS; UNMODIFIED KETONES; TERNARY COMPLEX

11709/CORDOVA A/2003/TETRAHEDRON LETTERS/DIRECT ORGANOCATALYTIC ASYMMETRIC MANNICH-TYPE REACTIONS IN AQUEOUS-MEDIA - ONE-POT MANNICH-ALLYLATION REACTIONS/NO FIELD/KEYWORDS PLUS: ALPHA-IMINO ESTERS; ALDOL REACTIONS; DIELS-ALDER; CATALYST; KETONES; COMPLEX

12554/CORDOVA A/2002/TETRAHEDRON LETTERS/ANTI-SELECTIVE SMP-CATALYZED DIRECT ASYMMETRIC MANNICH-TYPE REACTIONS - SYNTHESIS OF FUNCTIONALIZED AMINO-ACID DERIVATIVES/NO FIELD/KEYWORDS PLUS: ALPHA-IMINO ESTERS; ALDOL REACTIONS; KETONES; COMPLEX

CLUSTER 11

5691/WU XY/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/HIGHLY ENANTIOSELECTIVE EPOXIDATION OF ALPHA,BETA- UNSATURATED ESTERS BY CHIRAL DIOXIRANE/NO FIELD/KEYWORDS PLUS: CATALYTIC ASYMMETRIC EPOXIDATION; GENERATED IN-SITU; ALPHA,BETA-EPOXY ESTERS; UNFUNCTIONALIZED ALKENES; OPTICAL RESOLUTION; HYDROXY ESTERS; TRANS-OLEFINS; BETA- HYDROXY; KETONES; ACID

7916/SHU LH/2003/JOURNAL OF ORGANIC CHEMISTRY/AN IMPROVED SYNTHESIS OF A KETONE CATALYST FOR ASYMMETRIC EPOXIDATION OF OLEFINS/NO FIELD/KEYWORDS PLUS: GENERATED IN-SITU; HIGHLY ENANTIOSELECTIVE EPOXIDATION; CHIRAL KETONE; UNFUNCTIONALIZED ALKENES; TERMINAL OLEFINS; FLUORO KETONES; CIS-OLEFINS; DIOXIRANES; OXONE; EFFICIENCY

8761/ARMSTRONG A/2002/JOURNAL OF ORGANIC CHEMISTRY/ENANTIOSELECTIVE EPOXIDATION OF ALKENES CATALYZED BY 2- FLUORO-N-CARBETHOXYTROPINONE AND RELATED TROPINONE DERIVATIVES/NO FIELD/KEYWORDS PLUS: GENERATED IN-SITU; MEDIATED ASYMMETRIC EPOXIDATION; CHIRAL KETONE; UNFUNCTIONALIZED OLEFINS; ALPHA-FLUOROCYCLOHEXANONES; DIOXIRANES; ACID; OXONE(R); DIMETHYLDIOXIRANE; CYCLOHEXANONES

9194/BORTOLINI O/2002/JOURNAL OF ORGANIC CHEMISTRY/IMPROVED ENANTIOSELECTIVITY IN THE EPOXIDATION OF CINNAMIC ACID-DERIVATIVES WITH DIOXIRANES FROM KETO BILE-ACIDS/NO FIELD/KEYWORDS PLUS: CATALYTIC ASYMMETRIC EPOXIDATION; GENERATED IN-SITU; CHIRAL KETONES; UNFUNCTIONALIZED OLEFINS; DEHYDROCHOLIC ACID; ALKENES; OXIDATIONS; REACTIVITY; GEOMETRY; C-2

9617/TIAN HQ/2002/JOURNAL OF ORGANIC CHEMISTRY/DESIGNING NEW CHIRAL KETONE CATALYSTS - ASYMMETRIC EPOXIDATION OF CIS-OLEFINS AND TERMINAL OLEFINS/NO FIELD/KEYWORDS PLUS: HIGHLY ENANTIOSELECTIVE EPOXIDATION; GENERATED IN-SITU; HYDROGEN-PEROXIDE H2O2; UNFUNCTIONALIZED ALKENES; ABSOLUTE-CONFIGURATION; EPHEDRINE DERIVATIVES; KINETIC RESOLUTION; CONJUGATED DIENES; TRANSITION-STATE; PRIMARY OXIDANT

14233/MATSUMOTO K/2002/TETRAHEDRON LETTERS/CHIRAL KETONE-CATALYZED ASYMMETRIC EPOXIDATION OF OLEFINS WITH OXONE(R)/NO FIELD/KEYWORDS PLUS: GENERATED IN-SITU; UNFUNCTIONALIZED OLEFINS; DIOXIRANES; EFFICIENCY; ALKENES; IMINES

CLUSTER 12.

480/HAUSTEDT LO/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/THE TOTAL SYNTHESES OF PHORBOXAZOLES - NEW CLASSICS IN NATURAL-PRODUCT SYNTHESIS/AUTHOR KEYWORDS: ANTITUMOR AGENTS; MACROLIDES; PHORBOXAZOLES; SYNTHESIS DESIGN; TOTAL SYNTHESIS/KEYWORDS PLUS: PETASIS-FERRIER REARRANGEMENT; CHIRAL LEWIS-ACIDS; SPONGE PHORBAS SP; STEREOSELECTIVE SYNTHESIS; ABSOLUTE-CONFIGURATION; MARINE SPONGE; CONVERGENT SYNTHESIS; ASYMMETRIC-SYNTHESIS; ALDOL ADDITIONS; SIDE-CHAIN

742/GONZALEZ MA/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/A CONVERGENT TOTAL-SYNTHESIS OF PHORBOXAZOLE-A/AUTHOR KEYWORDS: ANTIFUNGAL AGENTS; ANTITUMOR AGENTS; NATURAL PRODUCTS; OLEFINATION; TOTAL SYNTHESIS/KEYWORDS PLUS: SPONGE PHORBAS SP; ABSOLUTE-CONFIGURATION; MARINE SPONGE; OLEFIN FORMATION; STEREOCHEMISTRY; MACROLIDE; SULFONES; ANALOGS; ESTERS

743/WILLIAMS DR/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/TOTAL-SYNTHESIS OF PHORBOXAZOLE-A/AUTHOR KEYWORDS: ANTITUMOR AGENTS; ASYMMETRIC ALLYLATION; MACROLIDES; NATURAL PRODUCTS; TOTAL SYNTHESIS/KEYWORDS PLUS: PETASIS-FERRIER REARRANGEMENT; SPONGE PHORBAS SP; STEREOSELECTIVE SYNTHESIS; CONVERGENT SYNTHESIS; MARINE SPONGE; ABSOLUTE-CONFIGURATION; ASYMMETRIC-SYNTHESIS; 1,3- DIOL ACETONIDES; NATURAL-PRODUCTS; SIDE-CHAIN

10188/LI DR/2003/TETRAHEDRON LETTERS/STUDIES ON THE SYNTHESIS OF PHORBOXAZOLE-B - STEREOSELECTIVE- SYNTHESIS OF THE C28-C46 SEGMENT/AUTHOR KEYWORDS: PHORBOXAZOLE B; SYNTHESIS; MUKAIYAMA ALDOL REACTION; OXAZOLE/KEYWORDS PLUS: PETASIS-FERRIER REARRANGEMENT; SPONGE PHORBAS SP; ASYMMETRIC-SYNTHESIS; CONVERGENT SYNTHESIS; NATURAL- PRODUCTS; SIDE-CHAIN; CONSTRUCTION; OXIDATION; FRAGMENT; SUBUNIT

11032/LIU B/2003/TETRAHEDRON LETTERS/STEREOSELECTIVE-SYNTHESIS OF THE C21-C27 FRAGMENT OF THE PHORBOXAZOLES/NO FIELD/KEYWORDS PLUS: PETASIS-FERRIER REARRANGEMENT; SPONGE PHORBAS SP; ASYMMETRIC-SYNTHESIS; SUBSTITUTED AZEPINONES; CONVERGENT SYNTHESIS; EFFICIENT SYNTHESIS; NATURAL-PRODUCTS; SIDE-CHAIN; SEGMENT; CONSTRUCTION

11244/PATERSON I/2003/TETRAHEDRON LETTERS/TOWARD THE TOTAL-SYNTHESIS OF PHORBOXAZOLE-A - SYNTHESIS OF AN ADVANCED C4-C32 SUBUNIT USING THE JACOBSEN HETERO-DIELS- ALDER REACTION/NO FIELD/KEYWORDS PLUS: PETASIS-FERRIER REARRANGEMENT; SPONGE PHORBAS SP; STEREOSELECTIVE SYNTHESIS; ALDOL REACTIONS; MARINE SPONGE; STEREOCONTROLLED SYNTHESIS; ABSOLUTE-CONFIGURATION; CONVERGENT SYNTHESIS; NATURAL-PRODUCTS; SIDE-CHAIN

CLUSTER 13

1397/TILLACK A/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/ANTI-MARKOVNIKOV HYDROAMINATION OF TERMINAL ALKYNES/AUTHOR KEYWORDS: ALKYNES; HOMOGENEOUS CATALYSIS; HYDROAMINATION; METALLOCENES; TITANIUM/KEYWORDS PLUS: CATALYZED INTERMOLECULAR HYDROAMINATION; OXIDATIVE AMINATION; AROMATIC OLEFINS; UNSATURATED-COMPOUNDS; H ACTIVATION; COMPLEXES; TITANOCENE; ALKENES; FUNCTIONALIZATIONS; DIMETHYLTITANOCENE

2480/ACKERMANN L/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/USE OF GROUP-4 BIS(SULFONAMIDO) COMPLEXES IN THE INTRAMOLECULAR HYDROAMINATION OF ALKYNES AND ALLENES/NO FIELD/KEYWORDS PLUS: CATALYZED INTERMOLECULAR HYDROAMINATION; N-H ACTIVATION; TERMINAL ALKYNES; ETA(1)-PYRROLYL COMPLEXES; SOLVENT PURIFICATION; IMIDO COMPLEXES; AMINES; CYCLIZATION; MECHANISM; SYSTEM

4996/SHIMADA T/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/PALLADIUM-CATALYZED INTERMOLECULAR HYDROAMINATION OF ALKYNES - A DRAMATIC RATE-ENHANCEMENT EFFECT OF O-AMINOPHENOL/NO FIELD/KEYWORDS PLUS: TERMINAL ALKYNES; COMPLEXES; DIMETHYLTITANOCENE; AMINATION; ALKENES; AMINES

9759/HEUTLING A/2002/JOURNAL OF ORGANIC CHEMISTRY/CP-ASTERISK-2TIME2 - AN IMPROVED CATALYST FOR THE INTERMOLECULAR ADDITION OF N-ALKYL-AMINE AND BENZYLAMINE TO ALKYNES/NO FIELD/KEYWORDS PLUS: ANTI-MARKOVNIKOV-FUNCTIONALIZATIONS; UNPROTECTED AMINO OLEFINS; 2+2 CYCLOADDITIONS; INTRAMOLECULAR HYDROAMINATION; REGIOSPECIFIC CYCLIZATION; SYNTHETIC APPLICATIONS; UNSATURATED-COMPOUNDS; OXIDATIVE AMINATION; TRANSITION-METALS; AROMATIC OLEFINS

13502/BYTSCHKOV T/2002/TETRAHEDRON LETTERS/THE CP(2)TIME2-CATALYZED INTRAMOLECULAR HYDROAMINATION/ CYCLIZATION OF AMINOALKYNES/AUTHOR KEYWORDS: ALKYNES; AMINATION; AMINES; CATALYSIS; TITANIUM/KEYWORDS PLUS: CATALYZED INTERMOLECULAR HYDROAMINATION; ALKYNE 2+2 CYCLOADDITIONS; SYNTHETIC APPLICATIONS; COMPLEXES; DIMETHYLTITANOCENE; CONCISE; AGENT

CLUSTER 14

1417/MOORE DR/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/ELECTRONIC AND STERIC EFFECTS ON CATALYSTS FOR CO2/EPOXIDE POLYMERIZATION - SUBTLE MODIFICATIONS RESULTING IN SUPERIOR ACTIVITIES/AUTHOR KEYWORDS: CARBON DIOXIDE FIXATION; GREEN CHEMISTRY; HOMOGENEOUS CATALYSIS; LIGAND EFFECTS; RING-OPENING POLYMERIZATION/KEYWORDS PLUS: CARBON-DIOXIDE; ALTERNATING COPOLYMERIZATION; CHROMIUM PORPHYRIN; CO2; EPOXIDES; PHENOXIDES; OXIDE; ZINC; DERIVATIVES; RELEVANCE

3252/DARENSBOURG DJ/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/COMPARATIVE KINETIC-STUDIES OF THE COPOLYMERIZATION OF CYCLOHEXENE OXIDE AND PROPYLENE-OXIDE WITH CARBON-DIOXIDE IN THE PRESENCE OF CHROMIUM SALEN DERIVATIVES - IN-SITU FTIR MEASUREMENTS OF COPOLYMER VS CYCLIC CARBONATE PRODUCTION/NO FIELD/KEYWORDS PLUS: ALTERNATING COPOLYMERIZATION; CATALYTIC ACTIVITY; POLYMER SYNTHESIS; EPOXIDES; CO2; COMPLEXES; PHENOXIDES; ZINC; INITIATION; REAGENTS

3645/NAKANO K/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/ASYMMETRIC ALTERNATING COPOLYMERIZATION OF CYCLOHEXENE OXIDE AND CO2 WITH DIMERIC ZINC-COMPLEXES/NO FIELD/KEYWORDS PLUS: CARBON-DIOXIDE; ENANTIOSELECTIVE ADDITION; PRECURSOR CATALYSTS; MECHANISTIC ASPECTS; CHROMIUM PORPHYRIN; MASS-SPECTROMETRY; POLYHYDRIC PHENOL; MAIN-CHAIN; POLYMERIZATION; EPOXIDES

4707/ALLEN SD/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/HIGH-ACTIVITY, SINGLE-SITE CATALYSTS FOR THE ALTERNATING COPOLYMERIZATION OF CO2 AND PROPYLENE-OXIDE/NO FIELD/KEYWORDS PLUS: CARBON-DIOXIDE; EPOXIDES; POLYMERIZATION; RELEVANCE; COMPLEXES; SYSTEM; ZINC

6107/DARENSBOURG DJ/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/MECHANISTIC ASPECTS OF THE COPOLYMERIZATION REACTION OF CARBON-DIOXIDE AND EPOXIDES, USING A CHIRAL SALEN CHROMIUM CHLORIDE CATALYST/NO FIELD/KEYWORDS PLUS: ALTERNATING COPOLYMERIZATION; CYCLOHEXENE OXIDE; POLYMER SYNTHESIS; COMPLEXES; CO2; PHENOXIDES; ZINC; FIXATION; SITES

CLUSTER 15

347/HAYASHI Y/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/THE DIRECT AND ENANTIOSELECTIVE, ONE-POT, 3-COMPONENT, CROSS- MANNICH REACTION OF ALDEHYDES/AUTHOR KEYWORDS: ALDEHYDES; ASYMMETRIC SYNTHESIS; ENANTIOSELECTIVITY; ORGANOCATALYSTS; 3-COMPONENT REACTION/KEYWORDS PLUS: AMINO-ACID-DERIVATIVES; ASYMMETRIC ALDOL REACTIONS; CHIRAL ZIRCONIUM CATALYST; DIELS-ALDER REACTION; ALPHA-IMINO ESTERS; BETA-AMINO; MICHAEL ADDITIONS; UNMODIFIED KETONES; ORGANIC CATALYSIS; ALLIBIS(BINAPHTHOXIDE) COMPLEX

704/JUHL K/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/THE 1ST ORGANOCATALYTIC ENANTIOSELECTIVE INVERSE-ELECTRON- DEMAND HETERO-DIELS-ALDER REACTION/NO FIELD/KEYWORDS PLUS: ASYMMETRIC ALPHA-AMINATION; 3-COMPONENT MANNICH REACTION; ORGANIC CATALYSIS; MICHAEL ADDITIONS; ALDOL REACTIONS; AMINO-ACIDS; 1,3-DIPOLAR CYCLOADDITION; CONJUGATE ADDITION; ALDEHYDES; KETONES

851/HALLAND N/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/HIGHLY ENANTIOSELECTIVE ORGANOCATALYTIC CONJUGATE ADDITION OF MALONATES TO ACYCLIC ALPHA,BETA-UNSATURATED ENONES/AUTHOR KEYWORDS: ASYMMETRIC CATALYSIS; ENONES; KETOESTERS; MALONATES; TETRAHYDROQUINOLINES/KEYWORDS PLUS: ASYMMETRIC MICHAEL ADDITION; BIS(OXAZOLINE) COPPER(II) COMPLEXES; ORGANIC CATALYSIS; ALDOL REACTIONS; L- PROLINE; STRATEGIES; KETONES; ACID; NITROALKANES

7933/MELCHIORRE P/2003/JOURNAL OF ORGANIC CHEMISTRY/DIRECT ENANTIOSELECTIVE MICHAEL ADDITION OF ALDEHYDES TO VINYL KETONES CATALYZED BY CHIRAL AMINES/NO FIELD/KEYWORDS PLUS: ASYMMETRIC CONJUGATE ADDITION; DIELS-ALDER REACTION; ORGANIC CATALYSIS; ALPHA-AMINATION; ALDOL REACTIONS; L-PROLINE; 1,3-DIPOLAR CYCLOADDITION; 1,4- CONJUGATE ADDITION; ACID; NITROALKANES

8726/HALLAND N/2002/JOURNAL OF ORGANIC CHEMISTRY/ORGANOCATALYTIC ASYMMETRIC CONJUGATE ADDITION OF NITROALKANES TO ALPHA,BETA-UNSATURATED ENONES USING NOVEL IMIDAZOLINE CATALYSTS/NO FIELD/KEYWORDS PLUS: PHASE-TRANSFER CATALYSTS; BOND-FORMING REACTIONS; MICHAEL ADDITIONS; ORGANIC CATALYSIS; CARBONYL-COMPOUNDS; ALDOL REACTIONS; KETONES; STRATEGIES; ALPHA

CLUSTER 16

2509/KIM MJ/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/(S)-SELECTIVE DYNAMIC KINETIC RESOLUTION OF SECONDARY ALCOHOLS BY THE COMBINATION OF SUBTILISIN AND AN AMINOCYCLOPENTADIENYLRUTHENIUM COMPLEX AS THE CATALYSTS/NO FIELD/KEYWORDS PLUS: ASYMMETRIC TRANSFORMATIONS; METAL CATALYSIS; CHIRAL ACETATES; EFFICIENT ROUTE; ENOL ACETATES; RACEMIZATION; AMINES; KETONES; ESTERS; ENZYME

8657/KIM MJ/2002/JOURNAL OF ORGANIC CHEMISTRY/ASYMMETRIC TRANSFORMATIONS OF ACYLOXYPHENYL KETONES BY ENZYME-METAL MULTICATALYSIS/NO FIELD/KEYWORDS PLUS: DYNAMIC KINETIC RESOLUTION; SECONDARY ALCOHOLS; HYDROGEN-TRANSFER; CHIRAL ACETATES; ENOL ACETATES; LIPASE; PALLADIUM; RACEMIZATION; ALDEHYDES; AMINES

10449/ITO M/2003/TETRAHEDRON LETTERS/RAPID RACEMIZATION OF CHIRAL NON-RACEMIC SEC-ALCOHOLS CATALYZED BY (ETA(5)-C-5(CH3)(5))RU COMPLEXES BEARING TERTIARY PHOSPHINE-PRIMARY AMINE CHELATE LIGANDS/NO FIELD/KEYWORDS PLUS: DYNAMIC KINETIC RESOLUTION; ASYMMETRIC HYDROGEN-TRANSFER; SECONDARY ALCOHOLS; ENZYMATIC RESOLUTION; ENOL ACETATES; RUTHENIUM; KETONES; MECHANISM; DIOLS; ROUTE

13672/RUNMO ABL/2002/TETRAHEDRON LETTERS/DYNAMIC KINETIC RESOLUTION OF GAMMA-HYDROXY ACID-DERIVATIVES/NO FIELD/KEYWORDS PLUS: RUTHENIUM-CATALYZED RACEMIZATION; CHIRAL BUILDING-BLOCKS; ENZYMATIC RESOLUTION; ENANTIOSELECTIVE SYNTHESIS; SECONDARY ALCOHOLS; ORGANIC-SOLVENTS; AMINO ALCOHOLS; LIPASE; LACTONES; COMPLEXES

CLUSTER 17

1763/HULTZSCH KC/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/THE 1ST POLYMER-SUPPORTED AND RECYCLABLE CHIRAL CATALYST FOR ENANTIOSELECTIVE OLEFIN METATHESIS/AUTHOR KEYWORDS: ASYMMETRIC CATALYSIS; IMMOBILIZATION; METATHESIS; MOLYBDENUM; SOLID-PHASE SYNTHESIS/KEYWORDS PLUS: RING-CLOSING METATHESIS; KINETIC RESOLUTION; COMPLEXES; KETONES; DERIVATIVES; ALKENES

2373/VANVELDHUIZEN JJ/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/CHIRAL RU-BASED COMPLEXES FOR ASYMMETRIC OLEFIN METATHESIS - ENHANCEMENT OF CATALYST ACTIVITY THROUGH STERIC AND ELECTRONIC MODIFICATIONS/NO FIELD/KEYWORDS PLUS: RING-CLOSING METATHESIS; OPENING-CROSS METATHESIS; ENANTIOSELECTIVE SYNTHESIS; ORGANIC-SYNTHESIS; UNSATURATED ALCOHOLS; RUTHENIUM CARBENE; TERTIARY ETHERS; STYRENYL ETHERS; EFFICIENT; ACRYLONITRILE

4114/TSANG WCP/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/AN ENANTIOMERICALLY PURE ADAMANTYLIMIDO MOLYBDENUM ALKYLIDENE COMPLEX - AN EFFECTIVE NEW CATALYST FOR ENANTIOSELECTIVE OLEFIN METATHESIS/NO FIELD/KEYWORDS PLUS: RING-CLOSING METATHESIS; CYCLIC TERTIARY ETHERS; LIGANDS; POLYMERIZATION

5370/TENG X/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/ENHANCEMENT OF ENANTIOSELECTIVITY BY THF IN ASYMMETRIC MO- CATALYZED OLEFIN METATHESIS. CATALYTIC ENANTIOSELECTIVE SYNTHESIS OF CYCLIC TERTIARY ETHERS AND SPIROCYCLES/NO FIELD/KEYWORDS PLUS: RING-CLOSING METATHESIS; IMIDO ALKYLIDENE COMPLEXES; CHIRAL ZIRCONIUM CATALYST; MANNICH-TYPE REACTIONS; KINETIC RESOLUTION; KETONES; IMINES; ACID

6005/DOLMAN SJ/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/EFFICIENT CATALYTIC ENANTIOSELECTIVE SYNTHESIS OF UNSATURATED AMINES - PREPARATION OF SMALL-RING AND MEDIUM- RING CYCLIC AMINES THROUGH MO-CATALYZED ASYMMETRIC RING- CLOSING METATHESIS IN THE ABSENCE OF SOLVENT/NO FIELD/KEYWORDS PLUS: OLEFIN METATHESIS; KINETIC RESOLUTION; STYRENYL ETHERS; HETEROCYCLES; CHROMENES; MECHANISM; ALKALOIDS; COMPLEXES

6288/VANVELDHUIZEN JJ/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/A RECYCLABLE CHIRAL RU CATALYST FOR ENANTIOSELECTIVE OLEFIN METATHESIS - EFFICIENT CATALYTIC ASYMMETRIC RING-OPENING/ CROSS METATHESIS IN AIR/NO FIELD/KEYWORDS PLUS: CROSS-METATHESIS; CLOSING METATHESIS; MECHANISM

6608/KIELY AF/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/ENANTIOSELECTIVE SYNTHESIS OF MEDIUM-RING HETEROCYCLES, TERTIARY ETHERS, AND TERTIARY ALCOHOLS BY MO-CATALYZED RING- CLOSING METATHESIS/NO FIELD/KEYWORDS PLUS: KINETIC RESOLUTION; KETONES

CLUSTER 18

1508/TAN DS/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/SYNTHESIS OF THE FUNCTIONALIZED TRICYCLIC SKELETON OF GUANACASTEPENE-A - A TANDEM EPOXIDE-OPENING BETA-ELIMINATION/ KNOEVENAGEL CYCLIZATION/NO FIELD/KEYWORDS PLUS: RING-SYSTEM; 3-OXO-4-PENTENOATE; CONVERSION; ALDEHYDES; FUNGUS

1509/LIN SN/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/A STEREOSELECTIVE ROUTE TO GUANACASTEPENE-A THROUGH A SURPRISING EPOXIDATION/NO FIELD/KEYWORDS PLUS: BIOLOGICAL-ACTIVITY; ALCOHOLS; CONVERSION; ALDEHYDES; REDUCTION; KETONES; FUNGUS; ALPHA

11164/MEHTA G/2003/TETRAHEDRON LETTERS/GUANACASTEPENE-A TOTAL-SYNTHESIS - CONSTRUCTION OF THE TRICYCLIC ISO-GUANACASTEPANE, EPI-GUANACASTEPANE AND GUANACASTEPANE FRAMEWORKS/NO FIELD/KEYWORDS PLUS: CARBON SKELETON; RING-SYSTEM; CORE; CYCLIZATION; PROGRESS; PORTION; FUNGUS

12141/MAGNUS P/2002/TETRAHEDRON LETTERS/SYNTHESIS OF THE HYDROAZULENE PORTION OF GUANACASTEPENE-A USING A (2.3)SIGMATROPIC SULFOXIDE REARRANGEMENT - OBSERVATIONS ON SILYL ENOL ETHER ELECTROPHILIC CHEMISTRY FOR THE INTRODUCTION OF THE C-13 HYDROXYL GROUP/AUTHOR KEYWORDS: GUANACASTEPENE; (2.3)SIGMATROPIC REARRANGEMENT; SILYL ENOL ETHERS/KEYWORDS PLUS: M-CHLOROPERBENZOIC ACID; STEREOSELECTIVE SYNTHESIS; ALPHA-HYDROXY; OXIDATION; FUNGUS

12579/BOYER FD/2002/TETRAHEDRON LETTERS/SYNTHESIS OF A HIGHLY FUNCTIONALIZED TRICYCLIC RING-SYSTEM RELATED TO GUANACASTEPENE VIA A TANDEM RING-CLOSING METATHESIS REACTION/NO FIELD/KEYWORDS PLUS: OLEFIN METATHESIS; DIENYNES; CONSTRUCTION; CATALYSTS; ALCOHOLS; EPOXIDES; FUNGUS

12714/MEHTA G/2002/TETRAHEDRON LETTERS/TOWARDS A TOTAL SYNTHESIS OF GUANACASTEPENE-A - CONSTRUCTION OF FULLY FUNCTIONALIZED AB AND BC RING SEGMENTS/NO FIELD/KEYWORDS PLUS: DOLASTANE DITERPENES; FUNGUS
13041/DUDLEY GB/2002/TETRAHEDRON LETTERS/ON THE USE OF DEUTERIUM-ISOTOPE EFFECTS IN CHEMICAL SYNTHESIS/NO FIELD/KEYWORDS PLUS: GUANACASTEPENE; FUNGUS

13407/NGUYEN TM/2002/TETRAHEDRON LETTERS/PROGRESS TOWARDS THE TOTAL SYNTHESIS OF GUANACASTEPENE-A - APPROACHES TO THE CONSTRUCTION OF QUATERNARY CARBONS AND THE 5-7-6-TRICYCLIC CARBON SKELETON/NO FIELD/KEYWORDS PLUS: C-H INSERTION; CYCLIZATION; FUNGUS

CLUSTER 19

9986/VOGL EM/2002/JOURNAL OF ORGANIC CHEMISTRY/PALLADIUM-CATALYZED MONOARYLATION OF NITROALKANES/NO FIELD/KEYWORDS PLUS: ALPHA-ARYLATION; KETONES; LIGANDS

110156/LIU P/2003/TETRAHEDRON LETTERS/A HIGHLY-ACTIVE CATALYST SYSTEM FOR THE HETEROARYLATION OF ACETONE/NO FIELD/KEYWORDS PLUS: ALPHA-ARYLATION; ASYMMETRIC ARYLATION; PALLADIUM; KETONES; ESTERS

13761/KASHIN AN/2002/TETRAHEDRON LETTERS/PALLADIUM-CATALYZED ARYLATION OF SULFONYL CH-ACIDS/AUTHOR KEYWORDS: PALLADIUM; SULFONES; CH-ACIDS; CATALYSIS; ARYL HALIDES; ARYLATION; CARBANIONS/KEYWORDS PLUS: ALPHA-ARYLATION; ARYL HALIDES; KETONES; COMPLEX

14370/TERAO Y/2002/TETRAHEDRON LETTERS/PALLADIUM-CATALYZED ALPHA-ARYLATION OF ALDEHYDES WITH ARYL BROMIDES/AUTHOR KEYWORDS: ARYL BALIDES; ARYLATION ALDEHYDES; PALLADIUM AND COMPOUNDS/KEYWORDS PLUS: ALPHA,BETA-UNSATURATED CARBONYL; REGIOSELECTIVE ARYLATION; KETONES; HALIDES; NAPHTHOLS

CLUSTER 20

11271/CHAN DMT/2003/TETRAHEDRON LETTERS/COPPER PROMOTED C-N AND C-O BOND CROSS-COUPLING WITH PHENYL AND PYRIDYLBORONATES/NO FIELD/KEYWORDS PLUS: ARYLBORONIC ACIDS; ROOM-TEMPERATURE; CUPRIC ACETATE; BORONIC ACIDS; DIARYL ETHERS; PHENYLBORONIC ACIDS; ARYLATION; ARYL; IMIDAZOLES; PHENOLS

11762/LAM PYS/2003/TETRAHEDRON LETTERS/N-ARYLATION OF ALPHA-AMINOESTERS WITH P-TOLYLBORONIC ACID- PROMOTED BY COPPER(II) ACETATE/NO FIELD/KEYWORDS PLUS: CROSS-COUPLING REACTIONS; ARYLBORONIC ACIDS; C-N; PHENYLBORONIC ACIDS; CUPRIC ACETATE; ROOM-TEMPERATURE; BORONIC ACIDS; DIARYL ETHERS; O-ARYLATION; AMINO-ACIDS

13699/LAM PYS/2002/TETRAHEDRON LETTERS/COPPER-PROMOTED C-N BOND CROSS-COUPLING WITH PHENYLSTANNANE/NO FIELD/KEYWORDS PLUS: ARYLBORONIC ACIDS; PHENYLBORONIC ACIDS; ROOM- TEMPERATURE; CUPRIC ACETATE; BORONIC ACIDS; DIARYL ETHERS; O-ARYLATION; PHENOLS; AMINES; IMIDAZOLES

CLUSTER 21

1376/KITA T/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/C-2-SYMMETRIC CHIRAL PENTACYCLIC GUANIDINE - A PHASE- TRANSFER CATALYST FOR THE ASYMMETRIC ALKYLATION OF TERT- BUTYL GLYCINATE SCHIFF-BASE/AUTHOR KEYWORDS: ALKYLATION; ASYMMETRIC SYNTHESIS; PHASE- TRANSFER CATALYSIS; SYNTHETIC METHODS/KEYWORDS PLUS: ALPHA-AMINO-ACIDS; QUATERNARY AMMONIUM SALT; SPONGE CRAMBE-CRAMBE; ENANTIOSELECTIVE SYNTHESIS; STEREOSELECTIVE SYNTHESIS; BICYCLIC GUANIDINE; MICHAEL REACTION; PTILOMYCALIN-A; AMIDINIUM IONS; ALKALOIDS

10175/ALLINGHAM MT/2003/TETRAHEDRON LETTERS/SYNTHESIS AND APPLICATIONS OF C-2-SYMMETRIC GUANIDINE BASES/AUTHOR KEYWORDS: PHASE TRANSFER; CATALYSIS; GUANIDINE/KEYWORDS PLUS: PHASE-TRANSFER CATALYSIS; ALPHA-AMINO-ACIDS; ENANTIOSELECTIVE SYNTHESIS; MICHAEL REACTION; CINCHONA ALKALOIDS; ALKYLATION; ESTERS

12125/ARAI S/2002/TETRAHEDRON LETTERS/PHASE-TRANSFER-CATALYZED ASYMMETRIC MICHAEL REACTION USING NEWLY-PREPARED CHIRAL QUATERNARY AMMONIUM-SALTS DERIVED FROM L-TARTRATE/AUTHOR KEYWORDS: ASYMMETRIC; CATALYSIS; MICHAEL REACTION; PHASE-TRANSFER/KEYWORDS PLUS: ALPHA-AMINO-ACIDS; ENANTIOSELECTIVE SYNTHESIS; LIGANDS

CLUSTER 22

779/DUTHALER RO/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/PROLINE-CATALYZED ASYMMETRIC ALPHA-AMINATION OF ALDEHYDES AND KETONES - AN ASTONISHINGLY SIMPLE ACCESS TO OPTICALLY- ACTIVE ALPHA-HYDRAZINO CARBONYL-COMPOUNDS/AUTHOR KEYWORDS: AMINATION; AZODICARBOXYLATES; CATALYSIS; ENANTIOSELECTIVITY; PROLINE/KEYWORDS PLUS: 3-COMPONENT MANNICH REACTION; ALDOL REACTIONS; AMINO-ACIDS; MICHAEL ADDITIONS; ENANTIOSELECTIVE AMINATION; ARYLGLYCINES; COMPLEXES; ALCOHOLS; ROUTE

3605/TANG Z/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/NOVEL SMALL ORGANIC-MOLECULES FOR A HIGHLY ENANTIOSELECTIVE DIRECT ALDOL REACTION/NO FIELD/KEYWORDS PLUS: ASYMMETRIC ALPHA-AMINATION; AMINO-ACIDS; ATOM ECONOMY; L-PROLINE; ALDEHYDES; CATALYSTS; KETONES; COMPLEX; ROUTE

4113/HARADA S/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/DIRECT CATALYTIC ASYMMETRIC MICHAEL REACTION OF HYDROXYKETONES - ASYMMETRIC ZN CATALYSIS WITH A ET2ZN/LINKED- BINOL COMPLEX/NO FIELD/KEYWORDS PLUS: LINKED-BINOL COMPLEX; MANNICH-TYPE REACTIONS; BIS(OXAZOLINE) COPPER(II) COMPLEXES; QUATERNARY AMMONIUM SALT; AMINO ACID-DERIVATIVES; SILYL ENOL ETHERS; ALDOL REACTION; UNMODIFIED KETONES; ALPHA-AMINO; ALLIBIS(BINAPHTHOXIDE) COMPLEX

CLUSTER 23

2207/ZHONG XH/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/ALLOYED ZNXCD1-XS NANOCRYSTALS WITH HIGHLY NARROW LUMINESCENCE SPECTRAL WIDTH/NO FIELD/KEYWORDS PLUS: LIGHT-EMITTING-DIODES; QUANTUM DOTS; SEMICONDUCTOR CLUSTERS; CDSE NANOCRYSTALS; SIZE; ZNS; NANOPARTICLES; CDXZN1-XS; POLYMER; ENERGY

2381/LI JJ/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/LARGE-SCALE SYNTHESIS OF NEARLY MONODISPERSE CDSE/CDS CORE/ SHELL NANOCRYSTALS USING AIR-STABLE REAGENTS VIA SUCCESSIVE ION LAYER ADSORPTION AND REACTION/NO FIELD/KEYWORDS PLUS: LIGHT-EMITTING-DIODES; SHELL QUANTUM DOTS; SEMICONDUCTOR NANOCRYSTALS; ALTERNATIVE ROUTES; EPITAXIAL- GROWTH; CLUSTERS; ZNSE; ELECTROLUMINESCENCE; NUCLEATION; DEPOSITION

3086/ZHONG XH/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/COMPOSITION-TUNABLE ZNXCD1-XSE NANOCRYSTALS WITH HIGH LUMINESCENCE AND STABILITY/NO FIELD/KEYWORDS PLUS: LIGHT-EMITTING-DIODES; QUANTUM DOTS; CORE/ SHELL NANOCRYSTALS; CDSE NANOCRYSTALS; EPITAXIAL-GROWTH; SIZE; CORE; SEMICONDUCTORS; NANOPARTICLES; CDXZN1-XS

3910/GUO WH/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/LUMINESCENT CDSE/CDS CORE/SHELL NANOCRYSTALS IN DENDRON BOXES - SUPERIOR CHEMICAL, PHOTOCHEMICAL AND THERMAL- STABILITY/NO FIELD/KEYWORDS PLUS: QUANTUM DOTS; SEMICONDUCTOR CLUSTERS; EPITAXIAL-GROWTH; CORED DENDRIMERS; MONODISPERSE; PHOTOLUMINESCENCE; NANOPARTICLES; COMPOSITES; PRECURSOR; NANORODS

6823/QU LH/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/CONTROL OF PHOTOLUMINESCENCE PROPERTIES OF CDSE NANOCRYSTALS IN GROWTH/NO FIELD/KEYWORDS PLUS: QUANTUM DOTS; SEMICONDUCTOR CLUSTERS; II-VI; EMISSION

CLUSTER 24

3357/STAUFFER SR/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/FLUORESCENCE RESONANCE ENERGY-TRANSFER (FRET) AS A HIGH- THROUGHPUT ASSAY FOR COUPLING REACTIONS - ARYLATION OF AMINES AS A CASE-STUDY/NO FIELD/KEYWORDS PLUS: PALLADIUM-CATALYZED AMINATION; ARYL CHLORIDES; HOMOGENEOUS CATALYSTS; MASS-SPECTROMETRY; ARYLBORONIC ACIDS; BOND FORMATION; C-N; ENANTIOSELECTIVE CATALYSTS; CROSS- COUPLINGS; MILD CONDITIONS

8536/URGAONKAR S/2003/JOURNAL OF ORGANIC CHEMISTRY/P(I-BUNCH2CH2)(3)N - AN EFFECTIVE LIGAND IN THE PALLADIUM- CATALYZED AMINATION OF ARYL BROMIDES AND IODIDES/NO FIELD/KEYWORDS PLUS: ROOM-TEMPERATURE AMINATION; PHOSPHORUS-NITROGEN BOND; CRYSTAL-STRUCTURE; SECONDARY-AMINES; CHLORIDES; HALIDES; SYSTEM; COMPLEXES; TRIS(MORPHOLINO)PHOSPHINE; PROAZAPHOSPHATRANES

8568/MARGOLIS BJ/2003/JOURNAL OF ORGANIC CHEMISTRY/AN EFFICIENT ASSEMBLY OF HETEROBENZAZEPINE RING-SYSTEMS UTILIZING AN INTRAMOLECULAR PALLADIUM-CATALYZED CYCLOAMINATION/NO FIELD/KEYWORDS PLUS: NITROGEN BOND FORMATION; ARYL HALIDES; AMINATION; CHLORIDES; BROMIDES; SCOPE

CLUSTER 25

1703/LEWIS FD/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/DYNAMICS AND ENERGETICS OF HOLE TRAPPING IN DNA BY 7- DEAZAGUANINE/AUTHOR KEYWORDS: DNA CONJUGATES; ELECTRON TRANSFER; PHOTOOXIDATION; PICOSECOND SPECTROSCOPY/KEYWORDS PLUS: PHOTOINDUCED ELECTRON-TRANSFER; DUPLEX DNA; TRANSPORT; OXIDATION; DISTANCE; SEQUENCE; HAIRPINS; 8- OXOGUANINE; DEPENDENCE; STACKING

2025/BELJONNE D/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/PATHWAYS FOR PHOTOINDUCED CHARGE SEPARATION IN DNA HAIRPINS/NO FIELD/KEYWORDS PLUS: ELECTRON-TRANSFER; HOLE TRANSPORT; DISTANCE DEPENDENCE; MIGRATION; MOLECULES; MECHANISM; PROTEINS; DYNAMICS; SEQUENCE; LINKERS

3739/LEWIS FD/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/DYNAMICS AND ENERGETICS OF SINGLE-STEP HOLE TRANSPORT IN DNA HAIRPINS/NO FIELD/KEYWORDS PLUS: DISTANCE CHARGE-TRANSPORT; PHOTOINDUCED ELECTRON-TRANSFER; DOUBLE-HELICAL DNA; B-FORM DNA; LONG- RANGE; TRANSIENT ABSORPTION; HOPPING MECHANISM; RADICAL-CATION; DEPENDENCE; OXIDATION

4779/LEWIS FD/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/FORMATION AND DECAY OF LOCALIZED CONTACT RADICAL-ION PAIRS IN DNA HAIRPINS/NO FIELD/KEYWORDS PLUS: PHOTOINDUCED ELECTRON-TRANSFER; CHARGE- TRANSFER; HOLE-TRANSPORT; DISTANCE DEPENDENCE; DYNAMICS; OLIGONUCLEOTIDES; DERIVATIVES; ENERGETICS; LINKERS; BASES

CLUSTER 26

1117/BEDFORD RB/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/SIMPLE MIXED TRICYCLOHEXYLPHOSPHANE-TRIARYLPHOSPHITE COMPLEXES AS EXTREMELY HIGH-ACTIVITY CATALYSTS FOR THE SUZUKI COUPLING OF ARYL CHLORIDES/AUTHOR KEYWORDS: C-C COUPLING; HOMOGENEOUS CATALYSIS; PALLADIUM; SUZUKI REACTION/KEYWORDS PLUS: ARYLBORONIC ACIDS; PHENYLBORONIC ACID; CROSS- COUPLINGS

1844/BOTELLA L/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/A CONVENIENT OXIME-CARBAPALLADACYCLE-CATALYZED SUZUKI CROSS- COUPLING OF ARYL CHLORIDES IN WATER/AUTHOR KEYWORDS: BIARYLS; C-C-COUPLING; CROSS-COUPLING; PALLADIUM; PHASE-TRANSFER CATALYSIS/KEYWORDS PLUS: N-HETEROCYCLIC CARBENES; HECK-TYPE REACTIONS; ARYLBORONIC ACIDS; PHENYLBORONIC ACID; COMPLEXES; HALIDES; PALLADACYCLES; PALLADIUM(II); CHLOROARENES; PRECURSORS

9166/ALONSO DA/2002/JOURNAL OF ORGANIC CHEMISTRY/HIGHLY-ACTIVE OXIME-DERIVED PALLADACYCLE COMPLEXES FOR SUZUKI-MIYAURA AND ULLMANN-TYPE COUPLING REACTIONS/NO FIELD/KEYWORDS PLUS: EFFICIENT CATALYST PRECURSORS; UNACTIVATED METHYL-GROUPS; ARYLBORONIC ACIDS; ARYL CHLORIDES; CROSS- COUPLINGS; C-C; STRUCTURAL CHARACTERIZATION; LIGANDLESS PALLADIUM; SYMMETRICAL BIARYLS; ROOM-TEMPERATURE

10348/TAO B/2003/TETRAHEDRON LETTERS/TRANS-PD(OAC)2(CY2NH)(2) CATALYZED SUZUKI COUPLING REACTIONS AND ITS TEMPERATURE-DEPENDENT ACTIVITIES TOWARD ARYL BROMIDES/NO FIELD/KEYWORDS PLUS: HIGHLY-ACTIVE CATALYSTS; N-HETEROCYCLIC CARBENES; ARYLBORONIC ACIDS; PALLADIUM CATALYSTS; C-C; CHLORIDES; COMPLEXES; WATER; CONVENIENT; LIGAND

13195/TAO B/2002/TETRAHEDRON LETTERS/PD(OAC)(2)/2-ARYL-2-OXAZOLINES CATALYZED SUZUKI COUPLING REACTIONS OF ARYL BROMIDES AND ARYLBORONIC ACIDS/NO FIELD/KEYWORDS PLUS: N-HETEROCYCLIC CARBENES; HIGHLY-ACTIVE CATALYSTS; HECK TYPE REACTIONS; EFFICIENT CATALYSTS; COMPLEXES; CHLORIDES; PALLADACYCLES; DERIVATIVES; CONVENIENT; VANCOMYCIN

CLUSTER 27

384/YAO QW/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/OLEFIN METATHESIS IN THE IONIC LIQUID 1-BUTYL-3- METHYLIMIDAZOLIUM HEXAFLUOROPHOSPHATE USING A RECYCLABLE RU CATALYST - REMARKABLE EFFECT OF A DESIGNER IONIC TAG/AUTHOR KEYWORDS: CARBENE COMPLEXES; HOMOGENEOUS CATALYSIS; IONIC LIQUIDS; METATHESIS; RUTHENIUM/KEYWORDS PLUS: ENYNE METATHESIS; EFFICIENT; COMPLEXES; LIGANDS; MECHANISM

1142/CONNON SJ/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/A SELF-GENERATING, HIGHLY-ACTIVE, AND RECYCLABLE OLERIN- METATHESIS CATALYST/AUTHOR KEYWORDS: ALKENES; HOMOGENEOUS CATALYSIS; METATHESIS; POLYMERS; RUTHENIUM/KEYWORDS PLUS: RING-CLOSING METATHESIS; OLEFIN-METATHESIS; IMIDAZOLIN-2-YLIDENE LIGANDS; MINIMAL PURIFICATION; CARBENE COMPLEXES; ORGANIC-SYNTHESIS; RUTHENIUM; POLYSTYRENE; EFFICIENT; PRODUCTS

12227/GRELA K/2002/TETRAHEDRON LETTERS/A PS-DES IMMOBILIZED RUTHENIUM CARBENE - A ROBUST AND EASILY RECYCLABLE CATALYST FOR OLEFIN METATHESIS/NO FIELD/KEYWORDS PLUS: CONVENIENT METHOD; ORGANIC-SYNTHESIS; CARBON- DIOXIDE; EFFICIENT; SUPPORTS

CLUSTER 28

879/CARDENAS DJ/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/ADVANCES IN FUNCTIONAL-GROUP-TOLERANT METAL-CATALYZED ALKYL- ALKYL CROSS-COUPLING REACTIONS/AUTHOR KEYWORDS: ALKANES; BORANES; CROSS-COUPLING; NICKEL; PALLADIUM/KEYWORDS PLUS: STILLE REACTION; ARYL CHLORIDES; EFFICIENT; MECHANISM; BROMIDES; CENTERS

1123/TSUJI T/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/COBALT-CATALYZED COUPLING REACTION OF ALKYL-HALIDES WITH ALLYLIC GRIGNARD-REAGENTS/AUTHOR KEYWORDS: ALLYLATION; C-C COUPLING; COBALT; CROSS- COUPLING; RADICAL REACTIONS/KEYWORDS PLUS: CARBON CENTERS; BETA-HYDROGENS; EFFICIENT; ALKENYLATION

1168/NETHERTON MR/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/SUZUKI CROSS-COUPLINGS OF ALKYL TOSYLATES THAT POSSESS BETA- HYDROGEN ATOMS - SYNTHETIC AND MECHANISTIC STUDIES/AUTHOR KEYWORDS: ALKYL TOSYLATES; C-C COUPLING; LIGAND EFFECTS; PALLADIUM; SUZUKI REACTION/KEYWORDS PLUS: BONDS

1539/KIRCHHOFF JH/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/A METHOD FOR PALLADIUM-CATALYZED CROSS-COUPLINGS OF SIMPLE ALKYL CHLORIDES - SUZUKI REACTIONS CATALYZED BY (PD- 2(DBA)(3))/PCY3/AUTHOR KEYWORDS: ALKYL CHLORIDES; CROSS-COUPLING; HOMOGENEOUS CATALYSIS; PALLADIUM; P LIGANDS/KEYWORDS PLUS: BETA-HYDROGENS; EFFICIENT; CENTERS

4813/KIRCHHOFF JH/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/BORONIC ACIDS - NEW COUPLING PARTNERS IN ROOM-TEMPERATURE SUZUKI REACTIONS OF ALKYL BROMIDES - CRYSTALLOGRAPHIC CHARACTERIZATION OF AN OXIDATIVE-ADDITION ADDUCT GENERATED UNDER REMARKABLY MILD CONDITIONS/NO FIELD/KEYWORDS PLUS: BETA-HYDROGENS

CLUSTER 29

9928/CARRIGAN MD/2002/JOURNAL OF ORGANIC CHEMISTRY/A SIMPLE AND EFFICIENT CHEMOSELECTIVE METHOD FOR THE CATALYTIC DEPROTECTION OF ACETALS AND KETALS USING BISMUTH TRIFLATE/NO FIELD/KEYWORDS PLUS: SELECTIVE CLEAVAGE; HIGHLY EFFICIENT; TRIFLUOROMETHANESULFONATE; CHLORIDE; ACYLATION

10364/PETERSON KE/2003/TETRAHEDRON LETTERS/BISMUTH COMPOUNDS IN ORGANIC-SYNTHESIS - SYNTHESIS OF RESORCINARENES USING BISMUTH TRIFLATE/AUTHOR KEYWORDS: BISMUTH AND COMPOUNDS; RESORCINARENES; LEWIS ACIDS; ENVIRONMENT-FRIENDLY CATALYSTS/KEYWORDS PLUS: COLUMNAR LIQUID-CRYSTALS; FRIEDEL-CRAFTS ACYLATION; HOST GUEST COMPLEXATION; DIELS-ALDER REACTION; EFFICIENT METHOD; ALDEHYDES; TRIFLUOROMETHANESULFONATE; CHLORIDE; KETONES; CALIX(4)RESORCINARENES

10703/REDDY AV/2003/TETRAHEDRON LETTERS/BISMUTH TRIFLATE CATALYZED CONJUGATE ADDITION OF INDOLES TO ALPHA,BETA-ENONES/AUTHOR KEYWORDS: BISMUTH TRIFLATE; INDOLE; ALPHA,BETA-ENONES; ADDITION REACTIONS/KEYWORDS PLUS: ELECTRON-DEFICIENT OLEFINS; DIELS-ALDER REACTION; EFFICIENT METHOD; HAPALOSIPHON-FONTINALIS; TRIFLUOROMETHANESULFONATE; HAPALINDOLES; ALDEHYDES; ALKALOIDS; CHLORIDE; KETONES

CLUSTER 30

672/CHOI TL/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/CONTROLLED LIVING RING-OPENING-METATHESIS POLYMERIZATION BY A FAST-INITIATING RUTHENIUM CATALYST/AUTHOR KEYWORDS: COPOLYMERIZATION; METATHESIS; N LIGANDS; RING-OPENING POLYMERIZATION; RUTHENIUM/KEYWORDS PLUS: OLEFIN CROSS-METATHESIS; N-HETEROCYCLIC CARBENES; ALKYLIDENE COMPLEXES; COPOLYMERS; LIGANDS; ROMP; NORBORNENES; MECHANISM

1089/LOVE JA/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/A PRACTICAL AND HIGHLY-ACTIVE RUTHENIUM-BASED CATALYST THAT EFFECTS THE CROSS METATHESIS OF ACRYLONITRILE/NO FIELD/KEYWORDS PLUS: OLEFIN METATHESIS; LIGANDS; EFFICIENT; VINYLPHOSPHONATE; MECHANISM; ALKENES; COMPLEX

1143/CHOI TL/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/SYNTHESIS OF A,B-ALTERNATING COPOLYMERS BY RING-OPENING- INSERTION-METATHESIS POLYMERIZATION/AUTHOR KEYWORDS: COPOLYMERIZATION; CROSS-COUPLING; METATHESIS; RING-OPENING POLYMERIZATION; RUTHENIUM/KEYWORDS PLUS: OLEFIN CROSS-METATHESIS; CATALYSTS; COMPLEXES; LIGANDS; ALKENES

1260/CHATTERJEE AK/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/FORMAL VINYL C-H ACTIVATION AND ALLYLIC OXIDATION BY OLEFIN METATHESIS/NO FIELD/KEYWORDS PLUS: CROSS-METATHESIS; CATALYSTS; ALKYNES

CLUSTER 31

1090/GRELA K/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/A HIGHLY EFFICIENT RUTHENIUM CATALYST FOR METATHESIS REACTION/NO FIELD/KEYWORDS PLUS: SELECTIVE CROSS-METATHESIS; OLEFIN METATHESIS; STABLE CARBENES; MECHANISM; BEARING; LIGANDS

1463/WAKAMATSU H/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/A NEW HIGHLY EFFICIENT RUTHENIUM METATHESIS CATALYST/NO FIELD/KEYWORDS PLUS: OLEFIN METATHESIS; CROSS METATHESIS; LIGANDS; COMPLEX

1734/WAKAMATSU H/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/A HIGHLY-ACTIVE AND AIR-STABLE RUTHENIUM COMPLEX FOR OLEFIN METATHESIS/AUTHOR KEYWORDS: BIARYLS; CARBENE LIGANDS; METATHESIS; O LIGANDS; RUTHENIUM/KEYWORDS PLUS: CATALYSTS

11514/DUNNE AM/2003/TETRAHEDRON LETTERS/A HIGHLY EFFICIENT OLEFIN METATHESIS INITIATOR - IMPROVED SYNTHESIS AND REACTIVITY STUDIES/NO FIELD/KEYWORDS PLUS: OPENING-CROSS-METATHESIS; IMIDAZOLIN-2- YLIDENE LIGANDS; RUTHENIUM COMPLEX; ORGANIC-SYNTHESIS; CATALYSTS; CYCLOALKENES; DERIVATIVES; GENERATION; CHEMISTRY

CLUSTER 32

5089/HAYASHI T/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/RHODIUM-CATALYZED ASYMMETRIC 1,4-ADDITION OF ARYLTITANIUM REAGENTS GENERATING CHIRAL TITANIUM ENOLATES - ISOLATION AS SILYL ENOL ETHERS/NO FIELD/KEYWORDS PLUS: CONJUGATE ADDITION; ARYLBORONIC ACIDS; ALPHA,BETA-UNSATURATED ESTERS; ORGANOBORONIC ACIDS; ALDOL REACTION; OLEFINS; KETONES; ENONES

5305/YOSHIDA K/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/A NEW-TYPE OF CATALYTIC TANDEM 1,4-ADDITION-ALDOL REACTION WHICH PROCEEDS THROUGH AN (OXA-PI-ALLYL)RHODIUM INTERMEDIATE/NO FIELD/KEYWORDS PLUS: REDUCTIVE ALDOL REACTION; ASYMMETRIC 1,4- ADDITION; CONJUGATE ADDITION; ARYLBORONIC ACIDS; ALPHA,BETA-UNSATURATED ESTERS; ORGANOBORONIC ACIDS; RHODIUM; REAGENTS; KETONES; ENONES

6309/HAYASHI T/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/CATALYTIC CYCLE OF RHODIUM-CATALYZED ASYMMETRIC 1,4-ADDITION OF ORGANOBORONIC ACIDS - ARYLRHODIUM, OXA-PI-ALLYLRHODIUM, AND HYDROXORHODIUM INTERMEDIATES/NO FIELD/KEYWORDS PLUS: ARYLBORONIC ACIDS; CONJUGATE ADDITION; ALPHA,BETA-UNSATURATED ESTERS; COUPLING REACTIONS; ENONES; REAGENTS; COMPLEXES; ALDEHYDES; ALDIMINES; LIGANDS

7743/ITOOKA R/2003/JOURNAL OF ORGANIC CHEMISTRY/RHODIUM-CATALYZED 1,4-ADDITION OF ARYLBORONIC ACIDS TO ALPHA,BETA-UNSATURATED CARBONYL-COMPOUNDS - LARGE ACCELERATING EFFECTS OF BASES AND LIGANDS/NO FIELD/KEYWORDS PLUS: ENANTIOSELECTIVE CONJUGATE ADDITION; ASYMMETRIC 1,4-ADDITION; ORGANOBORONIC ACIDS; GRIGNARD- REAGENTS; COUPLING REACTIONS; BASIC CONDITIONS; AQUEOUS- MEDIUM; CYCLIC ENONES; ALDEHYDES; COMPLEXES

8335/YOSHIDA K/2003/JOURNAL OF ORGANIC CHEMISTRY/GENERATION OF CHIRAL BORON ENOLATES BY RHODIUM-CATALYZED ASYMMETRIC 1,4-ADDITION OF 9-ARYL-9-BORABICYCLO(3.3.1)NONANES (B-AR-9BBN) TO ALPHA,BETA- UNSATURATED KETONES/NO FIELD/KEYWORDS PLUS: ARYLBORONIC ACIDS; ORGANOBORONIC ACIDS; CONJUGATE ADDITION; REAGENTS; ENONES; PHOSPHINE; LIGANDS; BINAP; CYCLOALKENONES; DERIVATIVES

CLUSTER 33

2516/HAYASHI T/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/A CHIRAL CHELATING DIENE AS A NEW-TYPE OF CHIRAL LIGAND FOR TRANSITION-METAL CATALYSTS - ITS PREPARATION AND USE FOR THE RHODIUM-CATALYZED ASYMMETRIC 1,4-ADDITION/NO FIELD/KEYWORDS PLUS: ARYLBORONIC ACIDS; CONJUGATE ADDITION; ALPHA,BETA-UNSATURATED KETONES; ORGANOBORONIC ACIDS; REAGENTS; ESTERS; ENONES; CYCLOALKENONES

4031/YOSHIDA K/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/A NEW CINE-SUBSTITUTION OF ALKENYL SULFONES WITH ARYLTITANIUM REAGENTS CATALYZED BY RHODIUM - MECHANISTIC STUDIES AND CATALYTIC ASYMMETRIC-SYNTHESIS OF ALLYLARENES/NO FIELD/KEYWORDS PLUS: ARYLBORONIC ACIDS; ORGANOBORONIC ACIDS; ALPHA,BETA-UNSATURATED ESTERS; CONJUGATE ADDITION; COUPLING REACTIONS; STILLE REACTION; 1,4-ADDITION; CYCLOALKENONES; COMPLEXES; LIGANDS

7237/BOITEAU JG/2003/JOURNAL OF ORGANIC CHEMISTRY/HIGH-EFFICIENCY AND ENANTIOSELECTIVITY IN THE RH-CATALYZED CONJUGATE ADDITION OF ARYLBORONIC ACIDS USING MONODENTATE PHOSPHORAMIDITES/NO FIELD/KEYWORDS PLUS: TANDEM 1,4-ADDITION-ALDOL REACTION; DIFFERENTIAL ACTIVATION ENTROPY; ASYMMETRIC 1,4-ADDITION; ORGANOBORONIC ACIDS; ALPHA,BETA-UNSATURATED ESTERS; KINETIC RESOLUTION; REAGENTS; LIGANDS; ENONES; CYCLOALKENONES

7570/DENMARK SE/2003/JOURNAL OF ORGANIC CHEMISTRY/PALLADIUM-CATALYZED CONJUGATE ADDITION OF ORGANOSILOXANES TO ALPHA,BETA-UNSATURATED CARBONYL-COMPOUNDS AND NITROALKENES/NO FIELD/KEYWORDS PLUS: CROSS-COUPLING REACTIONS; HYPERVALENT SILOXANE DERIVATIVES; ELECTRON-DEFICIENT OLEFINS; ASYMMETRIC 1,4-ADDITION; ARYLBORONIC ACIDS; ORGANOBORONIC ACIDS; ARYL HALIDES; PALLADIUM(0)-CATALYZED SILYLATION; BASIC CONDITIONS; SILVER(I) OXIDE

10681/SHI Q/2003/TETRAHEDRON LETTERS/BIPYRIDYL-BASED DIPHOSPHINE AS AN EFFICIENT LIGAND IN THE RHODIUM-CATALYZED ASYMMETRIC CONJUGATE ADDITION OF ARYLBORONIC ACIDS TO ALPHA,BETA-UNSATURATED KETONES/NO FIELD/KEYWORDS PLUS: CHIRAL DIPYRIDYLPHOSPHINE LIGAND; ORGANOBORONIC ACIDS; MICHAEL ADDITIONS; BETA-KETOESTERS; 1,4- ADDITION; HYDROGENATION; CYCLOALKENONES; NITROALKENES; REAGENTS; ENONES

CLUSTER 34

258/RAMACHARY DB/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/ ORGANOCATALYTIC ASYMMETRIC DOMINO KNOEVENAGEL/DIELS-ALDER REACTIONS - A BIOORGANIC APPROACH TO THE DIASTEREOSPECIFIC AND ENANTIOSELECTIVE CONSTRUCTION OF HIGHLY SUBSTITUTED SPIRO(5,5)UNDECANE-1,5,9-TRIONES/AUTHOR KEYWORDS: AMINO ACIDS; ASYMMETRIC CATALYSIS; CYCLOADDITION; DOMINO REACTIONS; ENANTIOSELECTIVITY/KEYWORDS PLUS: AMINO-ACID-DERIVATIVES; MANNICH-TYPE REACTIONS; ALPHA,BETA-UNSATURATED KETONES; ALPHA-AMINO; ORGANIC-SYNTHESIS; ALDOL REACTIONS; ALDEHYDES; CHEMISTRY; CATALYSTS; ROUTE

12760/RAMACHARY DB/2002/TETRAHEDRON LETTERS/AMINE-CATALYZED DIRECT SELF DIELS-ALDER REACTIONS OF ALPHA,BETA-UNSATURATED KETONES IN WATER - SYNTHESIS OF PRO-CHIRAL CYCLOHEXANONES/AUTHOR KEYWORDS: AMINES; CYCLOHEXANONES; DIELS-ALDER REACTIONS; ORGANOCATALYSIS; ENAMINES; IMINES; AQUEOUS MEDIA/KEYWORDS PLUS: ASYMMETRIC ALDOL REACTIONS; PROLINE

13439/THAYUMANAVAN R/2002/TETRAHEDRON LETTERS/AMINE-CATALYZED DIRECT DIELS-ALDER REACTIONS OF ALPHA,BETA- UNSATURATED KETONES WITH NITRO OLEFINS/AUTHOR KEYWORDS: AMINES; CATALYSIS; CYCLOHEXANONES; DIELS ALDER REACTIONS; ENAMINES; MICHAEL REACTIONS/KEYWORDS PLUS: ASYMMETRIC ALDOL REACTIONS; ENANTIOSELECTIVE SYNTHESIS; MICHAEL ADDITIONS; 4-NITROCYCLOHEXANONES; 2-AMINO- 1,3-BUTADIENES; CYCLOADDITION; NITROALKENES; DERIVATIVES

CLUSTER 35

714/GADEMANN K/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/THE 4TH HELICAL SECONDARY STRUCTURE OF BETA-PEPTIDES - THE (P)-2(8)-HELIX OF A BETA-HEXAPEPTIDE CONSISTING OF (2R,3S)-3- AMINO-2-HYDROXY ACID RESIDUES/AUTHOR KEYWORDS: AMINO ACIDS; BETA-PEPTIDES; CONFORMATIONAL ANALYSIS; PEPTIDOMIMETICS; SECONDARY STRUCTURE/KEYWORDS PLUS: AMINO-ACID; OLIGOMERS; FOLDAMERS; DESIGN

1552/MARTINEK TA/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/CIS-2-AMINOCYCLOPENTANECARBOXYLIC ACID OLIGOMERS ADOPT A SHEET-LIKE STRUCTURE - SWITCH FROM HELIX TO NONPOLAR STRAND/AUTHOR KEYWORDS: AMINO ACIDS; CHIRALITY; CONFORMATION ANALYSIS; NMR SPECTROSCOPY; PEPTIDES/KEYWORDS PLUS: BETA-PEPTIDES; SECONDARY STRUCTURE; SIDE- CHAINS; SPECTROSCOPY; TURNS

2121/SHARMA GVM/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/ROBUST MIXED 10/12 HELICES PROMOTED BY ALTERNATING CHIRALITY IN A NEW FAMILY OF C-LINKED CARBO-BETA-PEPTIDES/NO FIELD/KEYWORDS PLUS: SUGAR AMINO-ACIDS; SECONDARY STRUCTURE; OLIGOMERS; DESIGN; GLYCOBIOLOGY; FOLDAMERS; NMR

CLUSTER 36

1332/PARK HG/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/HIGHLY ENANTIOSELECTIVE AND PRACTICAL CINCHONA-DERIVED PHASE- TRANSFER CATALYSTS FOR THE SYNTHESIS OF ALPHA-AMINO-ACIDS/NO FIELD/KEYWORDS PLUS: QUATERNARY AMMONIUM-SALTS; ASYMMETRIC- SYNTHESIS; TRANSFER ALKYLATION; DERIVATIVES; IMINES; SOLVENT

3686/OOI T/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/DESIGN OF N-SPIRO C-2-SYMMETRIC CHIRAL QUATERNARY AMMONIUM BROMIDES AS NOVEL CHIRAL PHASE-TRANSFER CATALYSTS - SYNTHESIS AND APPLICATION TO PRACTICAL ASYMMETRIC-SYNTHESIS OF ALPHA-AMINO-ACIDS/NO FIELD/KEYWORDS PLUS: HIGHLY ENANTIOSELECTIVE SYNTHESIS; WEITZ- SCHEFFER EPOXIDATION; L-DOPA ESTERS; MICHAEL-ADDITION; STEREOSELECTIVE SYNTHESIS; D-GLUCOSE; CINCHONA ALKALOIDS; DARZENS REACTION; AZACROWN ETHERS; CROWN-ETHERS

11318/PARK HG/2003/TETRAHEDRON LETTERS/HIGHLY EFFICIENT ORTHO-FLUORO-DIMERIC CINCHONA-DERIVED PHASE- TRANSFER CATALYSTS/NO FIELD/KEYWORDS PLUS: ALPHA-AMINO-ACIDS; ENANTIOSELECTIVE SYNTHESIS; ASYMMETRIC-SYNTHESIS

12126/SHIBUGUCHI T/2002/TETRAHEDRON LETTERS/DEVELOPMENT OF NEW ASYMMETRIC 2-CENTER CATALYSTS IN PHASE- TRANSFER REACTIONS/AUTHOR KEYWORDS: PHASE-TRANSFER CATALYSIS; ASYMMETRIC 2- CENTER CATALYST; ALKYLATION; MICHAEL ADDITION/KEYWORDS PLUS: ALPHA-AMINO-ACIDS; ENANTIOSELECTIVE SYNTHESIS; ALKYLATION; DERIVATIVES; EPOXIDATION; IMINE

CLUSTER 37

17/HILLS ID/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/TOWARD AN IMPROVED UNDERSTANDING OF THE UNUSUAL REACTIVITY OF PD(0)/TRIALKYLPHOSPHANE CATALYSTS IN CROSS-COUPLINGS OF ALKYL ELECTROPHILES - QUANTIFYING THE FACTORS THAT DETERMINE THE RATE OF OXIDATIVE ADDITION/AUTHOR KEYWORDS: CROSS-COUPLING; HOMOGENEOUS CATALYSIS; PALLADIUM PHOSPHANE; LIGANDS; REACTION MECHANISMS/KEYWORDS PLUS: GRIGNARD-REAGENTS; ARYLBORONIC ACIDS; SUZUKI REACTIONS; MILD CONDITIONS; BETA-HYDROGENS; EFFICIENT; CHLORIDES; HALIDES; ARYL; DERIVATIVES

125/TANG HF/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/LIGANDS FOR PALLADIUM-CATALYZED CROSS-COUPLINGS OF ALKYL- HALIDES - USE OF AN ALKYLDIAMINOPHOSPHANE EXPANDS THE SCOPE OF THE STILLE REACTION/AUTHOR KEYWORDS: CROSS-COUPLING; HOMOGENEOUS CATALYSIS; PALLADIUM; PHOSPHANE LIGANDS; STILLE REACTION/KEYWORDS PLUS: GRIGNARD-REAGENTS; SUZUKI REACTIONS; BETA- HYDROGENS; BOND FORMATION; EFFICIENT; CHLORIDES; BROMIDES; DERIVATIVES; CENTERS; ACIDS

1936/ZHOU JR/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/CROSS-COUPLINGS OF UNACTIVATED SECONDARY ALKYL-HALIDES - ROOM-TEMPERATURE NICKEL-CATALYZED NEGISHI REACTIONS OF ALKYL BROMIDES AND IODIDES/NO FIELD/KEYWORDS PLUS: SUZUKI REACTIONS; BETA-HYDROGENS; EFFICIENT; TOSYLATES; CENTERS

2107/ECKHARDT M/2003/JOURNAL OF THE AMERICAN CHEMICAL
SOCIETY/THE 1ST APPLICATIONS OF CARBENE LIGANDS IN CROSS-COUPLINGS OF ALKYL ELECTROPHILES - SONOGASHIRA REACTIONS OF UNACTIVATED ALKYL BROMIDES AND IODIDES/NO FIELD/KEYWORDS PLUS: GRIGNARD-REAGENTS; SUZUKI REACTIONS; BETA-HYDROGENS; PALLADIUM(II); COMPLEXES; CHLORIDES; TOSYLATES; HALIDES

2376/ZHOU JR/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/PALLADIUM-CATALYZED NEGISHI CROSS-COUPLING REACTIONS OF UNACTIVATED ALKYL IODIDES, BROMIDES, CHLORIDES, AND TOSYLATES/NO FIELD/KEYWORDS PLUS: GRIGNARD-REAGENTS; SUZUKI REACTIONS; BETA- HYDROGENS; EFFICIENT; HALIDES; DERIVATIVES; SECONDARY; CENTERS

3542/LEE JY/2003/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/ROOM-TEMPERATURE HIYAMA CROSS-COUPLINGS OF ARYLSILANES WITH ALKYL BROMIDES AND IODIDES/NO FIELD/KEYWORDS PLUS: SUZUKI REACTIONS; CHLORIDES; BONDS

CLUSTER 38

722/MARIGO M/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/CATALYTIC, HIGHLY ENANTIOSELECTIVE, DIRECT AMINATION OF BETA- KETOESTERS/AUTHOR KEYWORDS: BETA-KETOESTERS; AMINATION; ASYMMETRIC CATALYSIS; COPPER; SYNTHETIC METHODS/KEYWORDS PLUS: AMINO-ACID-DERIVATIVES; ASYMMETRIC ALPHA- AMINATION; CHIRAL ZIRCONIUM CATALYST; MANNICH-TYPE REACTIONS; STRECKER REACTION; HYDROGEN-CYANIDE; IMINO ESTERS; COMPLEXES; ALDEHYDES; ALCOHOLS

1577/BOGEVIG A/2002/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/DIRECT ORGANO-CATALYTIC ASYMMETRIC ALPHA-AMINATION OF ALDEHYDES - A SIMPLE APPROACH TO OPTICALLY-ACTIVE ALPHA- AMINO ALDEHYDES, ALPHA-AMINO ALCOHOLS, AND ALPHA-AMINO-ACIDS/AUTHOR KEYWORDS: ALDEHYDES; AMINO ACIDS; AMINO ALCOHOLS; AMINO ALDEHYDES; ASYMMETRIC CATALYSIS/KEYWORDS PLUS: ENANTIOSELECTIVE SYNTHESIS; ZIRCONIUM CATALYST; STRECKER REACTION; HYDROGEN-CYANIDE; ALDOL REACTIONS; STRATEGIES; IMINES; ARYLGLYCINES; DERIVATIVES; ALKYLATION

6095/KUMARAGURUBARAN N/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/DIRECT L-PROLINE-CATALYZED ASYMMETRIC ALPHA-AMINATION OF KETONES/NO FIELD/KEYWORDS PLUS: AMINO-ACID DERIVATIVES; ORGANIC CATALYSIS; ENANTIOSELECTIVE AMINATION; ZIRCONIUM CATALYST; UNMODIFIED KETONES; STRECKER REACTION; HYDROGEN-CYANIDE; MANNICH REACTION; HYDRAZINO ACIDS; ALDOL REACTIONS

CLUSTER 39

830/REETZ MT/2003/ANGEWANDTE CHEMIE-INTERNATIONAL EDITION/A NEW PRINCIPLE IN COMBINATORIAL ASYMMETRIC TRANSITION-METAL CATALYSIS - MIXTURES OF CHIRAL MONODENTATE P LIGANDS/AUTHOR KEYWORDS: ASYMMETRIC CATALYSIS; COMBINATORIAL CHEMISTRY; HYDROGENATION; PHOSPHORUS; RHODIUM/KEYWORDS PLUS: ENANTIOSELECTIVE HYDROGENATION; ENAMIDES; DISCOVERY

4674/PENA D/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/HIGHLY ENANTIOSELECTIVE RHODIUM-CATALYZED HYDROGENATION OF BETA-DEHYDROAMINO ACID-DERIVATIVES USING MONODENTATE PHOSPHORAMIDITES/NO FIELD/KEYWORDS PLUS: AMINO ACIDS; ASYMMETRIC HYDROGENATION; LIGANDS

10962/NAKANO D/2003/TETRAHEDRON LETTERS/ENANTIOSELECTIVE HYDROGENATION OF ITACONATE USING RHODIUM BIHELICENOL PHOSPHITE COMPLEX - MATCHED/MISMATCHED PHENOMENA BETWEEN HELICAL AND AXIAL CHIRALITY/NO FIELD/KEYWORDS PLUS: ASYMMETRIC HYDROGENATION; CATALYZED HYDROGENATION; LIGANDS

12520/REETZ MT/2002/TETRAHEDRON LETTERS/ENANTIOSELECTIVE HYDROGENATION OF ENAMIDES CATALYZED BY CHIRAL RHODIUM-MONODENTATE PHOSPHITE COMPLEXES/NO FIELD/KEYWORDS PLUS: ASYMMETRIC HYDROGENATION; LIGANDS

CLUSTER 40

8552/BOSE DS/2003/JOURNAL OF ORGANIC CHEMISTRY/GREEN CHEMISTRY APPROACHES TO THE SYNTHESIS OF 5- ALKOXYCARBONYL-4-ARYL-3,4-DIHYDROPYRIMIDIN-2(1H)-ONES BY A 3-COMPONENT COUPLING OF ONE-POT CONDENSATION REACTION - COMPARISON OF ETHANOL, WATER, AND SOLVENT-FREE CONDITIONS/NO FIELD/KEYWORDS PLUS: BIGINELLI DIHYDROPYRIMIDINE SYNTHESIS; EFFICIENT SYNTHESIS; ORGANIC-SYNTHESIS; 3-COMPONENT; CATALYSIS; BLOCKERS; PROTOCOL

10315/REDDY KR/2003/TETRAHEDRON LETTERS/NEW ENVIRONMENTALLY FRIENDLY SOLVENT-FREE SYNTHESIS OF DIHYDROPYRIMIDINONES CATALYZED BY N-BUTYL-N,N-DIMETHYL-ALPHA- PHENYLETHYLAMMONIUM BROMIDE/AUTHOR KEYWORDS: N-BUTYL-N,N-DIMETHYL-ALPHA- PHENYLETHYLAMMONIUM BROMIDE; BETA-KETOESTERS; BIGINELLI REACTION; DIHYDROPYRIMIDINONES/KEYWORDS PLUS: CALCIUM-CHANNEL MODULATORS; ONE-POT SYNTHESIS; BIGINELLI REACTION; EFFICIENT SYNTHESIS; CONDENSATION REACTION; 3-COMPONENT; 3,4-DIHYDROPYRIMIDIN-2(1H)-ONES; REVISION; PROTOCOL; CHLORIDE

10774/TU S/2003/TETRAHEDRON LETTERS/ONE-POT SYNTHESIS OF 3,4-DIHYDROPYRIMIDIN-2(1H)-ONES USING BORIC-ACID AS CATALYST/NO FIELD/KEYWORDS PLUS: CALCIUM-CHANNEL BLOCKERS; BIGINELLI REACTION; ANTIHYPERTENSIVE AGENTS; MICROWAVE IRRADIATION; FLUOROUS SYNTHESIS; DIHYDROPYRIMIDINONES; ESTERS; CHLORIDE; 5- ALKOXYCARBONYL-4-ARYL-3,4-DIHYDROPYRIMIDIN-2(1H)-ONES; REVISION

11355/PARASKAR AS/2003/TETRAHEDRON LETTERS/CU(OTF)(2) - A REUSABLE CATALYST FOR HIGH-YIELD SYNTHESIS OF 3,4-DIHYDROPYRIMIDIN-2(1H)-ONES/AUTHOR KEYWORDS: ALDEHYDES; BIGINELLI REACTIONS; CATALYSTS; 3,4-DIHYDROPYRIMIDIN-2(IH)-ONES; COPPER AND COMPOUNDS/KEYWORDS PLUS: ONE-POT SYNTHESIS; BIGINELLI DIHYDROPYRIMIDINE SYNTHESIS; SOLVENT-FREE CONDITIONS; CONDENSATION REACTION; EFFICIENT SYNTHESIS; IMPROVED PROTOCOL; 3-COMPONENT; CHLORIDE

11458/SALEHI P/2003/TETRAHEDRON LETTERS/SILICA SULFURIC-ACID - AN EFFICIENT AND REUSABLE CATALYST FOR THE ONE-POT SYNTHESIS OF 3,4-DIHYDROPYRIMIDIN-2(1H)-ONES/AUTHOR KEYWORDS: BIGINELLI REACTION; SILICA SULFURIC ACID; DIHYDROPYRIMIDINONES; CATALYSIS; SOLID PHASE/KEYWORDS PLUS: BIGINELLI DIHYDROPYRIMIDINE SYNTHESIS; CALCIUM-CHANNEL BLOCKERS; HETEROGENEOUS SYSTEM; MILD CONDITIONS; CONDENSATION REACTION; ESTERS; 3-COMPONENT; ACID/ NANO2; PROTOCOL; MONASTROL

11520/MAITI G/2003/TETRAHEDRON LETTERS/ONE-POT SYNTHESIS OF DIHYDROPYRIMIDINONES CATALYZED BY LITHIUM BROMIDE - AN IMPROVED PROCEDURE FOR THE BIGINELLI REACTION/AUTHOR KEYWORDS: DIHYDROPYRIMIDINONES; LITHIUM BROMIDE; BIGINELLI REACTION/KEYWORDS PLUS: ALDEHYDES; REVISION

11968/SHAABANI A/2003/TETRAHEDRON LETTERS/AMMONIUM CHLORIDE-CATALYZED ONE-POT SYNTHESIS OF 3,4- DIHYDROPYRIMIDIN-2-(1H)-ONES UNDER SOLVENT-FREE CONDITIONS/AUTHOR KEYWORDS: BIGINELLI REACTION; DIHYDROPYRIMIDINONES; AMMONIUM CHLORIDE; ONE-POT CONDENSATION; SOLVENT-FREE/KEYWORDS PLUS: BIGINELLI REACTION; EFFICIENT SYNTHESIS; CONDENSATION REACTION; PARALLEL SYNTHESIS; DIHYDROPYRIMIDINONES; 3-COMPONENT; ACID; REVISION; PROTOCOL; ESTERS

CLUSTER 41

10046/HUANG JW/2003/TETRAHEDRON LETTERS/LEWIS-ACID BF3-CENTER-DOT-OET2-CATALYZED FRIEDEL-CRAFTS REACTION OF METHYLENECYCLOPROPANES WITH ARENES/AUTHOR KEYWORDS: METHYLENECYCLOPROPANES; LEWIS ACID; AROMATIC COMPOUNDS; FRIEDEL-CRAFTS REACTION; RING-OPENING REACTION/KEYWORDS PLUS: PALLADIUM; ALKYLIDENECYCLOPROPANES; (3+2)- CYCLOADDITION; PRONUCLEOPHILES

12455/SHI M/2002/TETRAHEDRON LETTERS/A NOVEL RING-OPENING REACTION OF METHYLENECYCLOPROPANES WITH AROMATIC-AMINES CATALYZED BY LEWIS-ACIDS/AUTHOR KEYWORDS: METHYLENECYCLOPROPANES (MCPS); AROMATIC AMINES; ALIPHATIC AMINES; RING-OPENING REACTIONS LEWIS ACIDS/KEYWORDS PLUS: PALLADIUM; ALKYLIDENECYCLOPROPANES; PRONUCLEOPHILES; (3+2)-CYCLOADDITION; HYDROCARBONATION; ALKENES

13726/XU B/2002/TETRAHEDRON LETTERS/THE REACTIONS OF THIOLS AND DIPHENYLDISULFIDE WITH TERMINALLY SUBSTITUTED METHYLENECYCLOPROPANES/AUTHOR KEYWORDS: THIOL; METHYLENECYCLOPROPANE; RADICAL REACTION/KEYWORDS PLUS: CATALYZED ADDITION; RADICAL CLOCKS; PALLADIUM; ACETYLENES

CLUSTER 42

5649/PASCALY M/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/DNA-MEDIATED CHARGE-TRANSPORT - CHARACTERIZATION OF A DNA RADICAL LOCALIZED AT AN ARTIFICIAL NUCLEIC-ACID BASE/NO FIELD/KEYWORDS PLUS: FLASH-QUENCH TECHNIQUE; RANGE ELECTRON- TRANSFER; DUPLEX DNA; HOLE TRANSPORT; HHAL METHYLTRANSFERASE; DISTANCE DEPENDENCE; FERROCYTOCHROME-C; HOPPING MECHANISM; OXIDATIVE DAMAGE; PULSE-RADIOLYSIS

6785/WILLIAMS TT/2002/JOURNAL OF THE AMERICAN CHEMICAL SOCIETY/THE EFFECT OF VARIED ION DISTRIBUTIONS ON LONG-RANGE DNA CHARGE-TRANSPORT/NO FIELD/KEYWORDS PLUS: ELECTRON-TRANSFER; DUPLEX DNA; OXIDATION; GUANINE; ASSEMBLIES; SEQUENCE; DAMAGE; FILMS; BASES

7599/DELANEY S/2003/JOURNAL OF ORGANIC CHEMISTRY/LONG-RANGE DNA CHARGE-TRANSPORT/NO FIELD/KEYWORDS PLUS: PHOTOINDUCED ELECTRON-TRANSFER; OXIDATIVE DAMAGE; CRYSTAL-STRUCTURE; GUANINE OXIDATION; LIGHT-SWITCH; DUPLEX DNA; DISTANCE; SEQUENCE; BASE; MECHANISM

CLUSTER 43

7205/CLIVE DLJ/2003/JOURNAL OF ORGANIC CHEMISTRY/DERIVATIZED AMINO-ACIDS RELEVANT TO NATIVE PEPTIDE-SYNTHESIS BY CHEMICAL LIGATION AND ACYL TRANSFER/NO FIELD/KEYWORDS PLUS: UNPROTECTED PEPTIDES; STAUDINGER LIGATION; PROTEIN-SYNTHESIS; SELENOCYSTEINE; SEGMENTS; ETHERS; AUXILIARY; REMOVAL; ESTERS; AZIDE

10752/KAWAKAMI T/2003/TETRAHEDRON LETTERS/A PHOTOREMOVABLE LIGATION AUXILIARY FOR USE IN POLYPEPTIDE- SYNTHESIS/NO FIELD/KEYWORDS PLUS: NATIVE CHEMICAL LIGATION; PROTEINS; THIOESTER

11095/MERKX R/2003/TETRAHEDRON LETTERS/CHEMOSELECTIVE COUPLING OF PEPTIDE-FRAGMENTS USING THE STAUDINGER LIGATION/AUTHOR KEYWORDS: AMIDE-FORMING LIGATION; AZIDO PEPTIDES; PEPTIDE O-(DIPHENYLPHOSPHINE)PHENYL ESTERS; STAUDINGER LIGATION/KEYWORDS PLUS: NATIVE CHEMICAL LIGATION; DIAZO TRANSFER; PROTEINS; AUXILIARY; AZIDE; ACIDS

CLUSTER 44

9712/BALAN D/2002/JOURNAL OF ORGANIC CHEMISTRY/TITANIUM ISOPROPOXIDE AS EFFICIENT CATALYST FOR THE AZA- BAYLIS-HILLMAN REACTION - SELECTIVE FORMATION OF ALPHA-METHYLENE-BETA-AMINO ACID-DERIVATIVES/NO FIELD/KEYWORDS PLUS: ACTIVATED DOUBLE-BONDS; TRICARBONYLCHROMIUM COMPLEXES; CARBONYL-COMPOUNDS; ESTERS; ALDEHYDES; ROUTE

11540/BALAN D/2003/TETRAHEDRON LETTERS/CHIRAL QUINUCLIDINE-BASED AMINE CATALYSTS FOR THE ASYMMETRIC ONE-POT, 3-COMPONENT AZA-BAYLIS-HILLMAN REACTION/NO FIELD/KEYWORDS PLUS: METHYL VINYL KETONE; ACID-DERIVATIVES; TRICARBONYLCHROMIUM COMPLEXES; SELECTIVE FORMATION; VERSION; IMINES

13849/CICLOSI M/2002/TETRAHEDRON LETTERS/SYNTHESIS OF UNSATURATED BETA-AMINO ACID-DERIVATIVES FROM CARBAMATES OF THE BAYLIS-HILLMAN PRODUCTS/NO FIELD/KEYWORDS PLUS: CONVENIENT APPROACH; PYRROLIDIN-2-ONES; BASICITY; CYCLIZATION; ALDEHYDES; SUPPORT; MN(III); ESTERS; DBU

APPENDIX 5

THE COMPARISON OF TWO PARTITIONS IN CASE 2

The Dispersion of Articles over Clusters for Two Partitions.

In the following table, columns A-D show the dispersion of articles in clusters generated by the field expert over the clusters generated by the complete link cluster method whereas columns E-H show the dispersion of articles in clusters generated by the complete link cluster method over the clusters generated by the field expert.

A	B	C	D	E	F	G	H
Complete Doc.nr.	Complete Clu.nr.	Expert Doc.nr.	Expert Clu.nr.	Complete Doc.nr.	Complete Clu.nr.	Expert Doc.nr.	Expert Clu.nr.
10297	4	10297	1	1014	1	1014	2
12144	4	12144	1	7888	1	7888	5
12579	18	12579	1	12254	1	12254	2
9986	19	9986	1	13307	1	13307	2
11762	20	11762	1	14376	1	14376	2
8568	24	8568	1	7415	2	7415	14
879	28	879	1	7732	2	7732	14
1143	30	1143	1	8686	2	8686	14
7743	32	7743	1	7672	3	7672	14
12760	34	12760	1	8098	3	8098	14
1936	37	1936	1	9348	3	9348	14
10774	40	10774	1	10297	4	10297	1
13849	44	13849	1	12144	4	12144	1
1014	1	1014	2	12650	4	12650	2
12254	1	12254	2	13490	4	13490	2
13307	1	13307	2	10480	4	10480	5
14376	1	14376	2	6342	5	6342	15
12650	4	12650	2	11268	5	11268	15
13490	4	13490	2	12470	5	12470	15
12402	8	12402	2	723	6	723	12
12575	8	12575	2	1112	6	1112	12
12730	8	12730	2	1579	6	1579	12
11709	10	11709	2	495	7	495	5
1397	13	1397	2	4105	7	4105	13
4996	13	4996	2	4502	7	4502	14
13502	13	13502	2	12402	8	12402	2
1508	18	1508	2	12575	8	12575	2
13761	19	13761	2	12730	8	12730	2
14370	19	14370	2	1640	9	1640	5
11271	20	11271	2	534	9	534	7
13699	20	13699	2	1631	9	1631	7
779	22	779	2	13651	9	13651	7
1844	26	1844	2	8856	9	8856	8
10348	26	10348	2	11709	10	11709	2
13195	26	13195	2	7018	10	7018	5
384	27	384	2	12554	10	12554	5
1123	28	1123	2	5691	11	5691	5
1168	28	1168	2	7916	11	7916	5

1539	28	1539	2	8761	11	8761	5
4813	28	4813	2	9194	11	9194	5
10703	29	10703	2	14233	11	14233	5
672	30	672	2	9617	11	9617	6
1260	30	1260	2	480	12	480	7
7570	33	7570	2	742	12	742	7
13439	34	13439	2	743	12	743	7
2107	37	2107	2	10188	12	10188	7
2376	37	2376	2	11032	12	11032	7
3542	37	3542	2	11244	12	11244	7
6095	38	6095	2	1397	13	1397	2
8552	40	8552	2	4996	13	4996	2
10315	40	10315	2	13502	13	13502	2
11520	40	11520	2	2480	13	2480	3
11968	40	11968	2	9759	13	9759	3
10046	41	10046	2	4707	14	4707	3
12455	41	12455	2	3645	14	3645	5
13726	41	13726	2	1417	14	1417	14
2480	13	2480	3	3252	14	3252	14
9759	13	9759	3	6107	14	6107	14
4707	14	4707	3	347	15	347	5
13041	18	13041	3	704	15	704	5
10156	19	10156	3	851	15	851	5
8536	24	8536	3	7933	15	7933	5
1117	26	1117	3	8726	15	8726	5
9166	26	9166	3	2509	16	2509	5
1142	27	1142	3	8657	16	8657	6
12227	27	12227	3	10449	16	10449	9
9928	29	9928	3	13672	16	13672	14
10364	29	10364	3	6608	17	6608	4
1089	30	1089	3	6005	17	6005	5
1090	31	1090	3	1763	17	1763	6
1463	31	1463	3	2373	17	2373	6
1734	31	1734	3	4114	17	4114	6
11514	31	11514	3	5370	17	5370	6
5305	32	5305	3	6288	17	6288	6
10681	33	10681	3	12579	18	12579	1
11318	36	11318	3	1508	18	1508	2
12126	36	12126	3	13041	18	13041	3
125	37	125	3	1509	18	1509	7
11355	40	11355	3	11164	18	11164	7
11458	40	11458	3	12141	18	12141	7
9712	44	9712	3	12714	18	12714	7
6608	17	6608	4	13407	18	13407	7
8335	32	8335	4	9986	19	9986	1
3686	36	3686	4	13761	19	13761	2
7888	1	7888	5	14370	19	14370	2
10480	4	10480	5	10156	19	10156	3
495	7	495	5	11762	20	11762	1
1640	9	1640	5	11271	20	11271	2
7018	10	7018	5	13699	20	13699	2

12554	10	12554	5	12125	21	12125	5
5691	11	5691	5	1376	21	1376	6
7916	11	7916	5	10175	21	10175	6
8761	11	8761	5	779	22	779	2
9194	11	9194	5	4113	22	4113	5
14233	11	14233	5	3605	22	3605	6
3645	14	3645	5	2207	23	2207	17
347	15	347	5	2381	23	2381	17
704	15	704	5	3086	23	3086	17
851	15	851	5	3910	23	3910	17
7933	15	7933	5	6823	23	6823	17
8726	15	8726	5	8568	24	8568	1
2509	16	2509	5	8536	24	8536	3
6005	17	6005	5	3357	24	3357	10
12125	21	12125	5	1703	25	1703	15
4113	22	4113	5	2025	25	2025	15
5089	32	5089	5	3739	25	3739	15
7237	33	7237	5	4779	25	4779	15
258	34	258	5	1844	26	1844	2
722	38	722	5	10348	26	10348	2
1577	38	1577	5	13195	26	13195	2
4674	39	4674	5	1117	26	1117	3
10962	39	10962	5	9166	26	9166	3
12520	39	12520	5	384	27	384	2
9617	11	9617	6	1142	27	1142	3
8657	16	8657	6	12227	27	12227	3
1763	17	1763	6	879	28	879	1
2373	17	2373	6	1123	28	1123	2
4114	17	4114	6	1168	28	1168	2
5370	17	5370	6	1539	28	1539	2
6288	17	6288	6	4813	28	4813	2
1376	21	1376	6	10703	29	10703	2
10175	21	10175	6	9928	29	9928	3
3605	22	3605	6	10364	29	10364	3
2516	33	2516	6	1143	30	1143	1
1332	36	1332	6	672	30	672	2
830	39	830	6	1260	30	1260	2
11540	44	11540	6	1089	30	1089	3
534	9	534	7	1090	31	1090	3
1631	9	1631	7	1463	31	1463	3
13651	9	13651	7	1734	31	1734	3
480	12	480	7	11514	31	11514	3
742	12	742	7	7743	32	7743	1
743	12	743	7	5305	32	5305	3
10188	12	10188	7	8335	32	8335	4
11032	12	11032	7	5089	32	5089	5
11244	12	11244	7	6309	32	6309	13
1509	18	1509	7	7570	33	7570	2
11164	18	11164	7	10681	33	10681	3
12141	18	12141	7	7237	33	7237	5
12714	18	12714	7	2516	33	2516	6

13407	18	13407	7	4031	33	4031	13
8856	9	8856	8	12760	34	12760	1
10449	16	10449	9	13439	34	13439	2
3357	24	3357	10	258	34	258	5
7205	43	7205	11	714	35	714	16
10752	43	10752	11	1552	35	1552	16
11095	43	11095	11	2121	35	2121	16
723	6	723	12	11318	36	11318	3
1112	6	1112	12	12126	36	12126	3
1579	6	1579	12	3686	36	3686	4
4105	7	4105	13	1332	36	1332	6
6309	32	6309	13	1936	37	1936	1
4031	33	4031	13	2107	37	2107	2
7415	2	7415	14	2376	37	2376	2
7732	2	7732	14	3542	37	3542	2
8686	2	8686	14	125	37	125	3
7672	3	7672	14	17	37	17	14
8098	3	8098	14	6095	38	6095	2
9348	3	9348	14	722	38	722	5
4502	7	4502	14	1577	38	1577	5
1417	14	1417	14	4674	39	4674	5
3252	14	3252	14	10962	39	10962	5
6107	14	6107	14	12520	39	12520	5
13672	16	13672	14	830	39	830	6
17	37	17	14	10774	40	10774	1
6342	5	6342	15	8552	40	8552	2
11268	5	11268	15	10315	40	10315	2
12470	5	12470	15	11520	40	11520	2
1703	25	1703	15	11968	40	11968	2
2025	25	2025	15	11355	40	11355	3
3739	25	3739	15	11458	40	11458	3
4779	25	4779	15	10046	41	10046	2
5649	42	5649	15	12455	41	12455	2
6785	42	6785	15	13726	41	13726	2
7599	42	7599	15	5649	42	5649	15
714	35	714	16	6785	42	6785	15
1552	35	1552	16	7599	42	7599	15
2121	35	2121	16	7205	43	7205	11
2207	23	2207	17	10752	43	10752	11
2381	23	2381	17	11095	43	11095	11
3086	23	3086	17	13849	44	13849	1
3910	23	3910	17	9712	44	9712	3
6823	23	6823	17	11540	44	11540	6

APPENDIX 6

BIBLIOGRAPHIC DESCRIPTIONS OF CORE DOCUMENT CLUSTERS IN CASE 2.

The bibliographic descriptions are delimited to document number and title.

CLUSTER 1

1703 DYNAMICS AND ENERGETICS OF HOLE TRAPPING IN DNA BY 7- DEAZAGUANINE

2746 DIRECT OBSERVATION OF GUANINE RADICAL-CATION DEPROTONATION IN DUPLEX DNA USING PULSE-RADIOLYSIS

3394 RAPID RADICAL FORMATION BY DNA CHARGE-TRANSPORT THROUGH SEQUENCES LACKING INTERVENING GUANINES

3498 BASE SEQUENCE EFFECTS IN RADICAL-CATION MIGRATION IN DUPLEX DNA - SUPPORT FOR THE POLARON-LIKE HOPPING MODEL

3676 RATIONAL DESIGN OF A DNA WIRE POSSESSING AN EXTREMELY HIGH HOLE TRANSPORT ABILITY

3739 DYNAMICS AND ENERGETICS OF SINGLE-STEP HOLE TRANSPORT IN DNA HAIRPINS

5965 N-2-PHENYLDEOXYGUANOSINE - MODULATION OF THE CHEMICAL- PROPERTIES OF DEOXYGUANOSINE TOWARD ONE-ELECTRON OXIDATION IN DNA

6342 DYNAMICS OF INTERSTRAND AND INTRASTRAND HOLE TRANSPORT IN DNA HAIRPINS

CLUSTER 2

17 TOWARD AN IMPROVED UNDERSTANDING OF THE UNUSUAL REACTIVITY OF PD(0)/TRIALKYLPHOSPHANE CATALYSTS IN CROSS-COUPLINGS OF ALKYL ELECTROPHILES– QUANTIFYING THE FACTORS THAT DETERMINE THE RATE OF OXIDATIVE ADDITION

125 LIGANDS FOR PALLADIUM-CATALYZED CROSS-COUPLINGS OF ALKYL- HALIDES - USE OF AN ALKYLDIAMINOPHOSPHANE EXPANDS THE SCOPE OF THE STILLE REACTION

1936 CROSS-COUPLINGS OF UNACTIVATED SECONDARY ALKYL-HALIDES-ROOM-TEMPERATURE NICKEL-CATALYZED NEGISHI REACTIONS OF ALKYL BROMIDES AND IODIDES

2376 PALLADIUM-CATALYZED NEGISHI CROSS-COUPLING REACTIONS OF UNACTIVATED ALKYLIODIDES, BROMIDES, CHLORIDES, AND TOSYLATES

4813 BORONIC ACIDS - NEW COUPLING PARTNERS IN ROOM-TEMPERATURE SUZUKI REACTIONS OF ALKYL BROMIDES - CRYSTALLOGRAPHIC CHARACTERIZATION OF AN OXIDATIVE-ADDITION ADDUCT GENERATED UNDER REMARKABLY MILD CONDITIONS

CLUSTER 3

1089 A PRACTICAL AND HIGHLY-ACTIVE RUTHENIUM-BASED CATALYST THAT EFFECTS THE CROSS METATHESIS OF ACRYLONITRILE

1142 A SELF-GENERATING, HIGHLY-ACTIVE, AND RECYCLABLE OLERIN- METATHESIS CATALYST

1463 A NEW HIGHLY EFFICIENT RUTHENIUM METATHESIS CATALYST

1734 A HIGHLY-ACTIVE AND AIR-STABLE RUTHENIUM COMPLEX FOR OLEFIN METATHESIS

12579 SYNTHESIS OF A HIGHLY FUNCTIONALIZED TRICYCLIC RING-SYSTEM RELATED TO GUANACASTEPENE VIA A TANDEM RING-CLOSING METATHESIS REACTION

CLUSTER 4

1014 CATALYTIC, ASYMMETRIC BAYLIS-HILLMAN REACTION OF IMINES WITH METHYL VINYLKETONE AND METHYL ACRYLATE

9712 TITANIUM ISOPROPOXIDE AS EFFICIENT CATALYST FOR THE AZA- BAYLIS-HILLMAN REACTION - SELECTIVE FORMATION OF ALPHA- METHYLENE-BETA-AMINO ACID-DERIVATIVES

12254 ONE-POT AZA-BAYLIS-HILLMAN REACTIONS OF ARYLALDEHYDES AND DIPHENYLPHOSPHINAMIDE WITH METHYL VINYL KETONE IN THE PRESENCE OF TICL4PPH3, AND ET3N

13307 BAYLIS-HILLMAN REACTIONS OF N- ARYLIDENEDIPHENYLPHOSPHINAMIDES WITH METHYL VINYL KETONE, METHYL ACRYLATE, AND ACRYLONITRILE

14376 LEWIS BASE AND L-PROLINE CO-CATALYZED BAYLIS-HILLMAN REACTION OF ARYLALDEHYDES WITH METHYL VINYL KETONE

CLUSTER 5

11355 CU(OTF)(2) - A REUSABLE CATALYST FOR HIGH-YIELD SYNTHESIS OF 3,4-DIHYDROPYRIMIDIN-2(1H)-ONES

11458 SILICA SULFURIC-ACID - AN EFFICIENT AND REUSABLE CATALYST FOR THE ONE-POT SYNTHESIS OF 3,4-DIHYDROPYRIMIDIN-2(1H)-ONES

CLUSTER 6

2516 A CHIRAL CHELATING DIENE AS A NEW-TYPE OF CHIRAL LIGAND FOR TRANSITION-METAL CATALYSTS - ITS PREPARATION AND USE FOR THE RHODIUM-CATALYZED ASYMMETRIC 1,4-ADDITION

4031 A NEW CINE-SUBSTITUTION OF ALKENYL SULFONES WITH ARYLTITANIUM REAGENTS CATALYZED BY RHODIUM - MECHANISTIC STUDIES AND CATALYTIC ASYMMETRIC-SYNTHESIS OF ALLYLARENES

5089 RHODIUM-CATALYZED ASYMMETRIC 1,4-ADDITION OF ARYLTITANIUM REAGENTS GENERATING CHIRAL TITANIUM ENOLATES - ISOLATION AS SILYL ENOL ETHERS

5305 A NEW-TYPE OF CATALYTIC TANDEM 1,4-ADDITION-ALDOL REACTION WHICH PROCEEDS THROUGH AN (OXA-PI-ALLYL)RHODIUM INTERMEDIATE

6309 CATALYTIC CYCLE OF RHODIUM-CATALYZED ASYMMETRIC 1,4-ADDITION OF ORGANOBORONIC ACIDS - ARYLRHODIUM, OXA-PI-ALLYLRHODIUM, AND HYDROXORHODIUM INTERMEDIATES

7743 RHODIUM-CATALYZED 1,4-ADDITION OF ARYLBORONIC ACIDS TO ALPHA,BETA-UNSATURATED CARBONYL-COMPOUNDS - LARGE ACCELERATING EFFECTS OF BASES AND LIGANDS

8335 GENERATION OF CHIRAL BORON ENOLATES BY RHODIUM-CATALYZED ASYMMETRIC 1,4-ADDITION OF 9-ARYL-9- BORABICYCLO(3.3.1)NONANES (B-AR-9BBN) TO ALPHA,BETA- NSATURATED KETONES

10681 BIPYRIDYL-BASED DIPHOSPHINE AS AN EFFICIENT LIGAND IN THE RHODIUM-CATALYZED ASYMMETRIC CONJUGATE ADDITION OF ARYLBORONIC ACIDS TO ALPHA,BETA-UNSATURATED KETONES

Cluster 7

262 A FACILE AND RAPID ROUTE TO HIGHLY ENANTIOPURE 1,2-DIOLS BY NOVEL CATALYTIC ASYMMETRIC ALPHA-AMINOXYLATION OF ALDEHYDES

347 THE DIRECT AND ENANTIOSELECTIVE, ONE-POT, 3-COMPONENT, CROSS- MANNICH REACTION OF ALDEHYDES

704 THE 1ST ORGANOCATALYTIC ENANTIOSELECTIVE INVERSE-ELECTRON- DEMAND HETERO-DIELS-ALDER REACTION

779 PROLINE-CATALYZED ASYMMETRIC ALPHA-AMINATION OF ALDEHYDES AND KETONES – AN ASTONISHINGLY SIMPLE ACCESS TO OPTICALLY- ACTIVE ALPHA-HYDRAZINOCARBONYL-COMPOUNDS

851 HIGHLY ENANTIOSELECTIVE ORGANOCATALYTIC CONJUGATE ADDITION OF MALONATES TO ACYCLIC ALPHA,BETA-UNSATURATED ENONES

3605 NOVEL SMALL ORGANIC-MOLECULES FOR A HIGHLY ENANTIOSELECTIVE DIRECT ALDOL REACTION

4113 DIRECT CATALYTIC ASYMMETRIC MICHAEL REACTION OF HYDROXYKETONES – ASYMMETRIC ZN CATALYSIS WITH A ET2ZN/LINKED- BINOL COMPLEX

4502 KINETIC AND STEREOCHEMICAL EVIDENCE FOR THE INVOLVEMENT OF ONLY ONE PROLINE MOLECULE IN THE TRANSITION-STATES OF PROLINE-CATALYZED INTRAMOLECULAR AND INTERMOLECULAR ALDOL REACTIONS

6095 DIRECT L-PROLINE-CATALYZED ASYMMETRIC ALPHA-AMINATION OF KETONES

7933 DIRECT ENANTIOSELECTIVE MICHAEL ADDITION OF ALDEHYDES TO VINYL KETONES CATALYZED BY CHIRAL AMINES

11273 PROLINE CATALYZED ALDOL REACTIONS IN AQUEOUS MICELLES - AN ENVIRONMENTALLY FRIENDLY REACTION SYSTEM

12554 ANTI-SELECTIVE SMP-CATALYZED DIRECT ASYMMETRIC MANNICH-TYPE REACTIONS – SYNTHESIS OF FUNCTIONALIZED AMINO-ACID DERIVATIVES

13439 AMINE-CATALYZED DIRECT DIELS-ALDER REACTIONS OF ALPHA,BETA- UNSATURATED KETONES WITH NITRO OLEFINS

APPENDIX 7

BIBLIOGRAPHIC DESCRIPTIONS OF CLUSTERS WITH A SIZE ≥ 3 IN CASE 3

Bibliographic data as follows: record number/ first author name/ publication year/ Journal name/ title/ author key words/key words plus. Missing data is indicated by "No Field".

CLUSTER 1

711/PATYI I/2003/MATHEMATISCHE ANNALEN/ON THE OKA PRINCIPLE IN A BANACH-SPACE, I/NO FIELD/NO FIELD

712/PATYI I/2003/MATHEMATISCHE ANNALEN/ON THE OKA PRINCIPLE IN A BANACH-SPACE, II/NO FIELD/NO FIELD

475/LEITERER J/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/A RELATIVE OKA-GRAUERT PRINCIPLE ON 1-CONVEX SPACES/NO FIELD/NO FIELD

CLUSTER 2

205/KARPENKO N/2003/INVENTIONES MATHEMATICAE/ESSENTIAL DIMENSION OF QUADRICS/NO FIELD/KEYWORDS PLUS: FIELDS

207/KARPENKO NA/2003/INVENTIONES MATHEMATICAE/ON THE 1ST WITT INDEX OF QUADRATIC-FORMS/NO FIELD/NO FIELD

463/MERKURJEV A/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/STEENROD OPERATIONS AND DEGREE FORMULAS/NO FIELD/NO FIELD

CLUSTER 3

454/GRIGORYAN A/2002/JOURNAL DE MATHEMATIQUES PURES ET APPLIQUEES/HITTING PROBABILITIES FOR BROWNIAN-MOTION ON RIEMANNIAN- MANIFOLDS/NO FIELD/KEYWORDS PLUS: HEAT KERNEL; MAXIMUM PRINCIPLE; DIFFUSION; BOUNDS

686/MURATA M/2003/MATHEMATISCHE ANNALEN/HEAT ESCAPE/NO FIELD/KEYWORDS PLUS: POSITIVE CAUCHY-PROBLEM; PARABOLIC HARNACK INEQUALITY; LOCAL DIRICHLET SPACES; INTRINSIC METRIC APPROACH; RIEMANNIAN-MANIFOLDS; ELLIPTIC-OPERATORS; SEMISMALL PERTURBATIONS; FUNDAMENTAL-SOLUTIONS; MARTIN BOUNDARIES; EQUATIONS

788/GRIGORYAN A/2002/MATHEMATISCHE ANNALEN/HARNACK INEQUALITIES AND SUB-GAUSSIAN ESTIMATES FOR RANDOM- WALKS/NO FIELD/KEYWORDS PLUS: BROWNIAN-MOTION; RIEMANNIAN-MANIFOLDS; SIERPINSKI CARPET; HEAT KERNEL; GRAPHS; NASH

CLUSTER 4

215/LEHN M/2003/INVENTIONES MATHEMATICAE/THE CUP PRODUCT OF HILBERT SCHEMES FOR K3 SURFACES/NO FIELD/KEYWORDS PLUS: COHOMOLOGY RING; POINTS; ALGEBRA; SHEAVES

559/LI WP/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/STABILITY OF THE COHOMOLOGY RINGS OF HILBERT SCHEMES OF POINTS ON SURFACES/NO FIELD/NO FIELD

665/LI WP/2002/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/ON BLOWUP FORMULAS FOR THE S-DUALITY CONJECTURE OF VAFA AND WITTEN III - RELATIONS WITH VERTEX OPERATOR-ALGEBRAS/NO FIELD/KEYWORDS PLUS: SURFACES; MONSTER; NUMBERS; SHEAVES; POINTS

803/LI WP/2002/MATHEMATISCHE ANNALEN/VERTEX ALGEBRAS AND THE COHOMOLOGY RING STRUCTURE OF HILBERT SCHEMES OF POINTS ON SURFACES/NO FIELD/KEYWORDS PLUS: VARIETY; PRODUCT

CLUSTER 5

240/LEARY IJ/2003/INVENTIONES MATHEMATICAE/SOME GROUPS OF TYPE VF/NO FIELD/KEYWORDS PLUS: FINITENESS PROPERTIES; ARTIN GROUPS; K-THEORY; CLASSIFYING SPACE; DISCRETE-GROUPS; ASSEMBLY MAPS; GRAPH GROUPS; FP-INFINITY; SUBGROUPS; CHARACTERS

279/LUCK W/2002/INVENTIONES MATHEMATICAE/THE RELATION BETWEEN THE BAUM-CONNES CONJECTURE AND THE TRACE CONJECTURE/NO FIELD/KEYWORDS PLUS: COUNTEREXAMPLE

656/LUCK W/2002/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/CHERN CHARACTERS FOR PROPER EQUIVARIANT HOMOLOGY THEORIES AND APPLICATIONS TO K-THEORY AND L-THEORY/NO FIELD/KEYWORDS PLUS: BAUM-CONNES CONJECTURE; DISCRETE-GROUPS; PROOF

CLUSTER 6

61/RAMAKRISHNA R/2002/ANNALS OF MATHEMATICS/DEFORMING GALOIS REPRESENTATIONS AND THE CONJECTURES OF SERRE AND FONTAINE-MAZUR/NO FIELD/KEYWORDS PLUS: MODULAR-REPRESENTATIONS; ELLIPTIC-CURVES; CONSTRUCTION; DEFORMATION; THEOREM; FORMS

192/KHARE C/2003/INVENTIONES MATHEMATICAE/FINITENESS OF SELMER GROUPS AND DEFORMATION RINGS/NO FIELD/KEYWORDS PLUS: GALOIS REPRESENTATIONS; EVEN REPRESENTATION

193/KHARE C/2003/INVENTIONES MATHEMATICAE/ON ISOMORPHISMS BETWEEN DEFORMATION RINGS AND HECKE RINGS/NO FIELD/KEYWORDS PLUS: MODULAR-REPRESENTATIONS; ELLIPTIC-CURVES; FORMS

CLUSTER 7

173/WANG BX/2002/COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS/THE LIMIT BEHAVIOR OF SOLUTIONS FOR THE CAUCHY-PROBLEM OF THE COMPLEX GINZBURG-LANDAU EQUATION/NO FIELD/KEYWORDS PLUS: NONLINEAR SCHRODINGER-EQUATIONS; DISPERSIVE EQUATIONS; KLEIN-GORDON; H-S; SCATTERING

798/MASMOUDI N/2002/MATHEMATISCHE ANNALEN/FROM NONLINEAR KLEIN-GORDON EQUATION TO A SYSTEM OF COUPLED NONLINEAR SCHRODINGER-EQUATIONS/NO FIELD/KEYWORDS PLUS: GLOBAL CAUCHY-PROBLEM; ENERGY SCATTERING

857/MACHIHARA S/2002/MATHEMATISCHE ANNALEN/NONRELATIVISTIC LIMIT IN THE ENERGY SPACE FOR NONLINEAR KLEIN-GORDON EQUATIONS/NO FIELD/KEYWORDS PLUS: GLOBAL CAUCHY-PROBLEM; SCHRODINGER-EQUATION

CLUSTER 8

220/KOBAYASHI S/2003/INVENTIONES MATHEMATICAE/IWASAWA THEORY FOR ELLIPTIC-CURVES AT SUPERSINGULAR PRIMES/NO FIELD/KEYWORDS PLUS: ABELIAN-VARIETIES; RATIONAL-POINTS; REDUCTION; FIELDS; TOWERS

281/KURIHARA M/2002/INVENTIONES MATHEMATICAE/ON THE TATE SHAFAREVICH GROUPS OVER CYCLOTOMIC FIELDS OF AN ELLIPTIC CURVE WITH SUPERSINGULAR REDUCTION I/NO FIELD/KEYWORDS PLUS: P-ADIC REPRESENTATIONS; IWASAWA THEORY; RATIONAL-POINTS; FORMAL GROUPS; CONJECTURE; BIRCH

499/KURIHARA M/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/IWASAWA THEORY AND FITTING IDEALS/NO FIELD/KEYWORDS PLUS: TOTALLY-REAL FIELDS; MINUS CLASS-GROUPS; ELLIPTIC-CURVES; ABELIAN-FIELDS; NUMBER-FIELDS; SELMER GROUPS; CONJECTURE; EXTENSIONS; RESIDUE; FORMULA

CLUSTER 9

437/REY O/2002/JOURNAL DE MATHEMATIQUES PURES ET APPLIQUEES/THE QUESTION OF INTERIOR BLOW-UP POINTS FOR AN ELLIPTIC NEUMANN PROBLEM - THE CRITICAL CASE/NO FIELD/KEYWORDS PLUS: CRITICAL SOBOLEV EXPONENTS; MULTI-PEAKED SOLUTIONS; LEAST-ENERGY SOLUTIONS; CAHN-HILLIARD EQUATION; CRITICAL NONLINEARITY; STATIONARY SOLUTIONS; EXISTENCE; BEHAVIOR; EQUILIBRIA; SYMMETRY

630/GUI CF/2002/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/ESTIMATES FOR BOUNDARY-BUBBLING SOLUTIONS TO AN ELLIPTIC NEUMANN PROBLEM/NO FIELD/KEYWORDS PLUS: CRITICAL SOBOLEV EXPONENTS; CRITICAL NONLINEARITY; EQUATIONS; BEHAVIOR

738/GROSSI M/2003/MATHEMATISCHE ANNALEN/A UNIQUENESS RESULT FOR A NEUMANN PROBLEM INVOLVING THE CRITICAL SOBOLEV EXPONENT/NO FIELD/NO FIELD

CLUSTER 10

99/LI AB/2003/COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS/ON SOME CONFORMALLY INVARIANT FULLY NONLINEAR EQUATIONS/NO FIELD/KEYWORDS PLUS: MONGE-AMPERE TYPE; ELLIPTIC-EQUATIONS; DIRICHLET PROBLEM; SCALAR CURVATURE; GEOMETRY; 4-MANIFOLDS; MANIFOLDS; SYMMETRY

494/SCHWETLICK H/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/CONVERGENCE OF THE YAMABE FLOW FOR LARGE ENERGIES/NO FIELD/KEYWORDS PLUS: CONFORMALLY FLAT MANIFOLDS; SCALAR CURVATURE; COMPACTNESS

535/GUAN PF/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/A FULLY NONLINEAR CONFORMAL FLOW ON LOCALLY CONFORMALLY FLAT MANIFOLDS/NO FIELD/KEYWORDS PLUS: POSITIVE RICCI CURVATURE; SCALAR CURVATURE; YAMABE FLOW; GEOMETRY; EQUATIONS

CLUSTER 11

269/GROVE K/2002/INVENTIONES MATHEMATICAE/COHOMOGENEITY ONE MANIFOLDS WITH POSITIVE RICCI CURVATURE/NO FIELD/KEYWORDS PLUS: HOMOGENEOUS EINSTEIN-METRICS; SPHERES; SPACES

295/WILKING B/2002/INVENTIONES MATHEMATICAE/MANIFOLDS WITH POSITIVE SECTIONAL CURVATURE ALMOST EVERYWHERE/NO FIELD/NO FIELD

334/BELEGRADEK I/2003/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/OBSTRUCTIONS TO NONNEGATIVE CURVATURE AND RATIONAL HOMOTOPY- THEORY/AUTHOR KEYWORDS: NONNEGATIVE CURVATURE; SOUL; DERIVATION; HALPERINS CONJECTURE/KEYWORDS PLUS: STRICTLY POSITIVE CURVATURE; HOMOGENEOUS SPACES; VECTOR-BUNDLES; MANIFOLDS; FINITENESS; SOUL

CLUSTER 12

84/FORNI G/2002/ANNALS OF MATHEMATICS/DEVIATION OF ERGODIC AVERAGES FOR AREA-PRESERVING FLOWS ON SURFACES OF HIGHER-GENUS/NO FIELD/KEYWORDS PLUS: INTERVAL EXCHANGE TRANSFORMATIONS; QUADRATIC- DIFFERENTIALS; MEASURED FOLIATIONS; TEICHMULLER-SPACES; DYNAMICAL-SYSTEMS; EXPONENTS; BILLIARDS; MANIFOLD

200/KONTSEVICH M/2003/INVENTIONES MATHEMATICAE/CONNECTED COMPONENTS OF THE MODULI SPACES OF ABELIAN DIFFERENTIALS WITH PRESCRIBED SINGULARITIES/NO FIELD/KEYWORDS PLUS: QUADRATIC-DIFFERENTIALS; TRANSFORMATIONS; FOLIATIONS

859/MUCINORAYMUNDO J/2002/MATHEMATISCHE ANNALEN/COMPLEX STRUCTURES ADAPTED TO SMOOTH VECTOR-FIELDS/NO FIELD/KEYWORDS PLUS: QUADRATIC-DIFFERENTIALS; PRESCRIBED SINGULARITIES; FLOWS

CLUSTER 13

242/ESNAULT H/2003/INVENTIONES MATHEMATICAE/VARIETIES OVER A FINITE-FIELD WITH TRIVIAL CHOW GROUP OF 0- CYCLES HAVE A RATIONAL POINT/NO FIELD/KEYWORDS PLUS: RIGID COHOMOLOGY; HODGE TYPE

628/CHIARELLOTTO B/2002/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/A COMPARISON THEOREM FOR WEIGHTS/NO FIELD/KEYWORDS PLUS: UNIPOTENT F-ISOCRYSTALS; RIGID COHOMOLOGY; PURITY

869/BESSER A/2002/MATHEMATISCHE ANNALEN/COLEMAN INTEGRATION USING THE TANNAKIAN FORMALISM/NO FIELD/KEYWORDS PLUS: UNIPOTENT F-ISOCRYSTALS; RIGID COHOMOLOGY

CLUSTER 14

363/EMERTON M/2002/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/SUPERSINGULAR ELLIPTIC-CURVES, THETA-SERIES AND WEIGHT 2 MODULAR-FORMS/NO FIELD/KEYWORDS PLUS: HECKE OPERATORS; MOD-P; REPRESENTATIONS

667/HALBERSTADT E/2002/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/FERMAT-CURVES - RESULTS AND PROBLEMS/NO FIELD/KEYWORDS PLUS: ELLIPTIC-CURVES; MODULAR-REPRESENTATIONS; LAST THEOREM; PRIME

677/EMERTON M/2003/MATHEMATISCHE ANNALEN/OPTIMAL QUOTIENTS OF MODULAR JACOBIANS/NO FIELD/KEYWORDS PLUS: ELLIPTIC-CURVES; HECKE OPERATORS; REPRESENTATIONS; FORMS; PARAMETRIZATIONS

CLUSTER 15

635/BERTIN MA/2002/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/ON THE REGULARITY OF VARIETIES HAVING AN EXTREMAL SECANT LINE/NO FIELD/KEYWORDS PLUS: CASTELNUOVO

804/GIRALDO L/2002/MATHEMATISCHE ANNALEN/ON THE PROJECTIVE-NORMALITY OF ENRIQUES SURFACES (WITH AN APPENDIX BY LOPEZ,ANGELO,FELICE AND VERRA,ALESSANDRO)/NO FIELD/KEYWORDS PLUS: VECTOR-BUNDLES; ALGEBRAIC-SURFACES; KOSZUL COHOMOLOGY; LINEAR-SYSTEMS; CURVE; SYZYGIES; VARIETIES; DIMENSION

871/NOMA A/2002/MATHEMATISCHE ANNALEN/A BOUND ON THE CASTELNUOVO-MUMFORD REGULARITY FOR CURVES/NO FIELD/KEYWORDS PLUS: SPACE

CLUSTER 16

261/FANG FQ/2002/INVENTIONES MATHEMATICAE/THE 2ND TWISTED BETTI NUMBER AND THE CONVERGENCE OF COLLAPSING RIEMANNIAN-MANIFOLDS/NO FIELD/KEYWORDS PLUS: POSITIVE SECTIONAL CURVATURE; DIAMETER; FINITENESS; GEOMETRY; HOMOTOPY

486/ANDERSON MT/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/SCALAR CURVATURE AND THE EXISTENCE OF GEOMETRIC STRUCTURES ON 3-MANIFOLDS, II/NO FIELD/KEYWORDS PLUS: COLLAPSING RIEMANNIAN-MANIFOLDS; VACUUM EINSTEIN EQUATIONS; METRIC DEGENERATIONS

564/ANDERSON MT/2002/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/SCALAR CURVATURE AND THE EXISTENCE OF GEOMETRIC STRUCTURES ON 3-MANIFOLDS, I/NO FIELD/KEYWORDS PLUS: COLLAPSING RIEMANNIAN-MANIFOLDS; VACUUM EINSTEIN EQUATIONS; METRIC DEGENERATIONS

867/TUSCHMANN W/2002/MATHEMATISCHE ANNALEN/GEOMETRIC DIFFEOMORPHISM FINITENESS IN LOW DIMENSIONS AND HOMOTOPY GROUP FINITENESS/NO FIELD/KEYWORDS PLUS: COLLAPSING RIEMANNIAN-MANIFOLDS; CONTROLLED TOPOLOGY; BOUNDING HOMOTOPY; CURVATURE; THEOREMS; DIAMETER

CLUSTER 17

12/KASPAROV G/2003/ANNALS OF MATHEMATICS/GROUPS ACTING PROPERLY ON BOLIC SPACES AND THE NOVIKOV- CONJECTURE/NO FIELD/KEYWORDS PLUS: BAUM-CONNES CONJECTURE; EQUIVARIANT KK-THEORY; ALGEBRAS

312/LAFFORGUE V/2002/INVENTIONES MATHEMATICAE/BIVARIANT K-THEORY FOR BANACH-ALGEBRAS AND BAUM-CONNES CONJECTURE/NO FIELD/KEYWORDS PLUS: CROSSED-PRODUCTS; TREES; PROOF; AMENABILITY; PROPERTY

472/EMERSON H/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/NONCOMMUTATIVE POINCARE-DUALITY FOR BOUNDARY ACTIONS OF HYPERBOLIC GROUPS/NO FIELD/KEYWORDS PLUS: NOVIKOV-CONJECTURE; KK-THEORY; ALGEBRAS

CLUSTER 18

18/JIANG DH/2003/ANNALS OF MATHEMATICS/THE LOCAL CONVERSE THEOREM FOR SO(2N+1) AND APPLICATIONS/NO FIELD/KEYWORDS PLUS: P-ADIC FIELD; RANKIN-SELBERG CONVOLUTIONS; MULTIPLICITY ONE THEOREM; GENERIC REPRESENTATIONS; THETA- LIFT; GL(N); CONJECTURE; MODULES; PROOF

21/LAPID E/2003/ANNALS OF MATHEMATICS/ON THE NONNEGATIVITY OF L(1/2, PI) FOR SO2N+1/NO FIELD/KEYWORDS PLUS: AUTOMORPHIC L-FUNCTIONS; IRREDUCIBLE REPRESENTATIONS; PLANCHEREL MEASURES; ADIC GROUPS; REDUCIBILITY; GL(N); CLASSIFICATION; CONJECTURE; VALUES; FIELD

72/KIM HH/2002/ANNALS OF MATHEMATICS/FUNCTORIAL PRODUCTS FOR GL(2)XGL(3) AND THE SYMMETRIC CUBE FOR GL(2)/NO FIELD/KEYWORDS PLUS: RANKIN-SELBERG CONVOLUTIONS; LANGLANDS- SHAHIDI METHOD; AUTOMORPHIC L-FUNCTIONS; CUSP FORMS; INTERTWINING-OPERATORS; FOURIER COEFFICIENTS; PLANCHEREL MEASURES; REPRESENTATIONS; CONJECTURE; GL(N)

73/BUSHNELL CJ/2002/ANNALS OF MATHEMATICS/APPENDIX - ON CERTAIN DYADIC REPRESENTATIONS/NO FIELD/KEYWORDS PLUS: RANKIN-SELBERG CONVOLUTIONS; GL(N)

349/KIM HH/2003/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/FUNCTORIALITY FOR THE EXTERIOR SQUARE OF GL(4) AND THE SYMMETRIC 4TH OF GL(2)/NO FIELD/KEYWORDS PLUS: P-ADIC FIELD; LOCAL LANGLANDS CONJECTURE; AUTOMORPHIC L-FUNCTIONS; INTERTWINING-OPERATORS; PLANCHEREL MEASURES; EULER PRODUCTS; SHAHIDI METHOD; CUSP FORMS; GL(N); REPRESENTATIONS

364/MOEGLIN C/2002/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/CONSTRUCTION OF DISCRETE-SERIES FOR CLASSICAL P-ADIC GROUPS/AUTHOR KEYWORDS: CLASSICAL GROUPS; P-ADIC FIELDS; IRREDUCIBLE SQUARE INTEGRABLE REPRESENTATIONS; IRREDUCIBLE TEMPERED REPRESENTATIONS; NONUNITARY DUAL; LOCAL LANGLANDS CORRESPONDENCES/KEYWORDS PLUS: INDUCED REPRESENTATIONS; INTERTWINING-OPERATORS; PLANCHEREL MEASURES; REDUCIBILITY; GL(N); NORMALIZATION; CONJECTURE; INDUCTION; PROOF

CLUSTER 19

523/HUYBRECHTS D/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/ FINITENESS RESULTS FOR COMPACT HYPERKAHLER MANIFOLDS/NO FIELD/KEYWORDS PLUS: KAHLER-MANIFOLDS

642/MARKMAN E/2002/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/GENERATORS OF THE COHOMOLOGY RING OF MODULI SPACES OF SHEAVES ON SYMPLECTIC SURFACES/NO FIELD/KEYWORDS PLUS: PROJECTIVE VARIETY; HILBERT SCHEME; EQUATIONS; MANIFOLDS

716/HUYBRECHTS D/2003/MATHEMATISCHE ANNALEN/THE KAHLER CONE OF A COMPACT HYPERKAHLER MANIFOLD/NO FIELD/NO FIELD

779/NAMIKAWA Y/2002/MATHEMATISCHE ANNALEN/COUNTER-EXAMPLE TO GLOBAL TORELLI PROBLEM FOR IRREDUCIBLE SYMPLECTIC-MANIFOLDS/NO FIELD/NO FIELD

CLUSTER 20

276/JONSSON M/2002/INVENTIONES MATHEMATICAE/STABLE MANIFOLDS OF HOLOMORPHIC DIFFEOMORPHISMS/NO FIELD/NO FIELD

426/DINH TC/2003/JOURNAL DE MATHEMATIQUES PURES ET APPLIQUEES/DYNAMICS OF POLYNOMIAL-LIKE MAPPINGS/NO FIELD/KEYWORDS PLUS: ENTROPY; CURRENTS; DIFFEOMORPHISMS; EXPONENTS; MAP

875/DINH TC/2002/MATHEMATISCHE ANNALEN/PERMUTABLE HOLOMORPHIC ENDOMORPHISMS OF P-K/NO FIELD/KEYWORDS PLUS: DYNAMICS

CLUSTER 21

94/KOHN RV/2003/COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS/UPPER BOUND ON THE COARSENING RATE FOR AN EPITAXIAL-GROWTH MODEL/NO FIELD/KEYWORDS PLUS: MOLECULAR-BEAM EPITAXY; SLOPE SELECTION; PHASE-TRANSITIONS; THIN-FILMS; CONTINUUM MODEL; GRADIENT THEORY; DYNAMICS; ENERGY; COMPACTNESS; DESORPTION

124/AMBROSIO L/2003/COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS/A VISCOSITY PROPERTY OF MINIMIZING MICROMAGNETIC CONFIGURATIONS/NO FIELD/KEYWORDS PLUS: ENERGY; COMPACTNESS; FIELDS

150/DESIMONE A/2002/COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS/A REDUCED THEORY FOR THIN-FILM MICROMAGNETICS/NO FIELD/KEYWORDS PLUS: ENERGY; FERROMAGNETISM; COMPACTNESS

162/CONTI S/2002/COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS/A GAMMA-CONVERGENCE RESULT FOR THE 2-GRADIENT THEORY OF PHASE-TRANSITIONS/NO FIELD/KEYWORDS PLUS: NONCONVEX VARIATIONAL-PROBLEMS; MINIMAL INTERFACE CRITERION; SINGULAR PERTURBATIONS; LOCAL MINIMIZERS; ENERGY; FERROELASTICS; MIXTURES; FIELDS

379/DELELLIS C/2003/JOURNAL DE MATHEMATIQUES PURES ET APPLIQUEES/THE RECTIFIABILITY OF ENTROPY MEASURES IN ONE SPACE DIMENSION/AUTHOR KEYWORDS: CONSERVATION LAWS; ENTROPY SOLUTIONS; SHOCKS; CONCENTRATION/KEYWORDS PLUS: ENERGY; MICROMAGNETICS; COMPACTNESS; REGULARITY

CLUSTER 22

8/CHEUNG Y/2003/ANNALS OF MATHEMATICS/HAUSDORFF DIMENSION OF THE SET OF NONERGODIC DIRECTIONS/NO FIELD/KEYWORDS PLUS: FOLIATIONS

319/MCMULLEN CT/2003/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/BILLIARDS AND TEICHMULLER CURVES ON HILBERT MODULAR SURFACES/NO FIELD/KEYWORDS PLUS: ARITHMETIC FUCHSIAN-GROUPS; QUADRATIC- DIFFERENTIALS; TRIANGULAR BILLIARDS; EMBEDDINGS; SPACES

580/MINSKY Y/2002/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/ NONDIVERGENCE OF HOROCYCLIC FLOWS ON MODULI SPACE/NO FIELD/KEYWORDS PLUS: INTERVAL EXCHANGE TRANSFORMATIONS; HOMOGENEOUS SPACES; TEICHMULLER SPACE; FOLIATIONS; MANIFOLDS; SURFACES; MAPS; SET

CLUSTER 23

309/BERNDTSSON B/2002/INVENTIONES MATHEMATICAE/THE PARTIAL-DERIVATIVE-EQUATION ON A POSITIVE CURRENT/NO FIELD/KEYWORDS PLUS: THEOREM

693/CHEN BY/2003/MATHEMATISCHE ANNALEN/THE BERGMAN METRIC ON COMPLETE KAHLER-MANIFOLDS/NO FIELD/KEYWORDS PLUS: PSEUDOCONVEX DOMAINS; THEOREM; KERNEL

830/MCNEAL JD/2002/MATHEMATISCHE ANNALEN/L-2 HARMONIC FORMS ON SOME COMPLETE KAHLER-MANIFOLDS/NO FIELD/KEYWORDS PLUS: PSEUDOCONVEX DOMAINS; BERGMAN-KERNEL; CONVEX DOMAINS; COHOMOLOGY; METRICS

CLUSTER 24

15/BRENDLE S/2003/ANNALS OF MATHEMATICS/GLOBAL EXISTENCE AND CONVERGENCE FOR A HIGHER-ORDER FLOW IN CONFORMAL GEOMETRY/NO FIELD/KEYWORDS PLUS: ZETA-FUNCTION DETERMINANTS; RICCI FLOW; 4- MANIFOLDS; CURVATURE; INVARIANT; EQUATION; METRICS; EXTREMALS; SURFACES

69/CHANG SYA/2002/ANNALS OF MATHEMATICS/AN EQUATION OF MONGE-AMPERE TYPE IN CONFORMAL GEOMETRY, AND 4-MANIFOLDS OF POSITIVE RICCI CURVATURE/NO FIELD/KEYWORDS PLUS: 2ND-ORDER ELLIPTIC-EQUATIONS; COMPACT RIEMANNIAN-MANIFOLDS; ZETA-FUNCTION DETERMINANTS; DIRICHLET PROBLEM; CRITICAL EXPONENT; SCALAR CURVATURE; YAMABE FLOW; 4- MANIFOLDS; INEQUALITY; REGULARITY

108/CHANG SYA/2003/COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS/THE INEQUALITY OF MOSER AND TRUDINGER AND APPLICATIONS TO CONFORMAL GEOMETRY/NO FIELD/KEYWORDS PLUS: KAHLER-EINSTEIN METRICS; PRESCRIBING GAUSSIAN CURVATURE; ZETA-FUNCTIONAL DETERMINANTS; SIMONS HIGGS-MODEL; SCALAR CURVATURE; SOBOLEV INEQUALITIES; RIEMANNIAN-MANIFOLDS; EXTREMAL METRICS; EXISTENCE; 4-MANIFOLDS

787/BRENDLE S/2002/MATHEMATISCHE ANNALEN/CURVATURE FLOWS ON SURFACES WITH BOUNDARY/NO FIELD/KEYWORDS PLUS: RICCI

CLUSTER 25

154/GUZZETTI D/2002/COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS/THE ELLIPTIC REPRESENTATION OF THE GENERAL PAINLEVE-VI EQUATION/NO FIELD/KEYWORDS PLUS: ORDINARY DIFFERENTIAL-EQUATIONS; 2- DIMENSIONAL ISING-MODEL; RATIONAL COEFFICIENTS; QUANTUM COHOMOLOGY; MONODROMY; DEFORMATION; TC

549/HERTLING C/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/TT-ASTERISK GEOMETRY, FROBENIUS MANIFOLDS, THEIR CONNECTIONS, AND THE CONSTRUCTION FOR SINGULARITIES/NO FIELD/KEYWORDS PLUS: MIXED HODGE-STRUCTURES; KAHLER-MANIFOLDS; PERIOD; MONODROMY

713/CAO HD/2003/MATHEMATISCHE ANNALEN/ON QUASI-ISOMORPHIC DGBV ALGEBRAS/NO FIELD/KEYWORDS PLUS: FROBENIUS MANIFOLD STRUCTURE; COHOMOLOGY; GRAVITY; SPACE

CLUSTER 26

212/POPA M/2003/INVENTIONES MATHEMATICAE/STABLE MAPS AND QUOT SCHEMES/NO FIELD/KEYWORDS PLUS: VECTOR-BUNDLES; FLAG VARIETIES; MODULI; CURVES; SURFACES

243/GIVENTAL A/2003/INVENTIONES MATHEMATICAE/QUANTUM K-THEORY ON FLAG MANIFOLDS, FINITE-DIFFERENCE TODA- LATTICES AND QUANTUM GROUPS/NO FIELD/KEYWORDS PLUS: COHOMOLOGY

321/BUCH AS/2003/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/GROMOV-WITTEN INVARIANTS ON GRASSMANNIANS/AUTHOR KEYWORDS: GROMOV-WITTEN INVARIANTS; GRASSMANNIANS; FLAG VARIETIES; SCHUBERT VARIETIES; QUANTUM COHOMOLOGY; LITTLEWOOD-RICHARDSON RULE/KEYWORDS PLUS: SCHUBERT POLYNOMIALS; FLAG MANIFOLDS; FUSION RULES; FORMULA

338/RIETSCH K/2003/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/TOTALLY POSITIVE TOEPLITZ MATRICES AND QUANTUM COHOMOLOGY OF PARTIAL FLAG VARIETIES/AUTHOR KEYWORDS: FLAG VARIETIES; QUANTUM COHOMOLOGY; TOTAL POSITIVITY/KEYWORDS PLUS: SCHUBERT POLYNOMIALS; PARAMETRIZATIONS; MANIFOLDS; RINGS

CLUSTER 27

79/GRODAL J/2002/ANNALS OF MATHEMATICS/HIGHER LIMITS VIA SUBGROUP COMPLEXES/NO FIELD/KEYWORDS PLUS: SPORADIC SIMPLE-GROUPS; COMPACT LIE-GROUPS; CLASSIFYING-SPACES; FINITE-GROUPS; MODULAR-REPRESENTATIONS; MACKEY FUNCTORS; P-GROUP; HOMOTOPY; COHOMOLOGY; HOMOLOGY

230/BROTO C/2003/INVENTIONES MATHEMATICAE/HOMOTOPY-EQUIVALENCES OF P-COMPLETED CLASSIFYING-SPACES OF FINITE-GROUPS/NO FIELD/NO FIELD

318/BROTO C/2003/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/THE HOMOTOPY-THEORY OF FUSION SYSTEMS/AUTHOR KEYWORDS: CLASSIFYING SPACE; P-COMPLETION; FINITE GROUPS; FUSION/KEYWORDS PLUS: CLASSIFYING-SPACES; HIGHER LIMITS; DECOMPOSITION; EXTENSIONS; SUBGROUPS; DIAGRAMS; CATEGORY; MODULES; MAPS; RING

CLUSTER 28

11/MOSHER L/2003/ANNALS OF MATHEMATICS/QUASI-ACTIONS ON TREES I - BOUNDED VALENCE/NO FIELD/KEYWORDS PLUS: BAUMSLAG-SOLITAR GROUPS; ISOMETRIC RIGIDITY; GEOMETRY

264/BONK M/2002/INVENTIONES MATHEMATICAE/QUASI-SYMMETRIC PARAMETRIZATIONS OF 2-DIMENSIONAL METRIC SPHERES/NO FIELD/KEYWORDS PLUS: CIRCLE PACKINGS; GOOD PARAMETERIZATIONS; HYPERBOLIC BUILDINGS; MEASURE-SPACES; UNIFORMIZATION; CONSTANT; THEOREM; MAPS

571/BOURDON M/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/LP-COHOMOLOGY AND BESOV-SPACES/NO FIELD/KEYWORDS PLUS: METRIC MEASURE-SPACES; HYPERBOLIC BUILDINGS; BOUNDARY

CLUSTER 29

48/SAPIR MV/2002/ANNALS OF MATHEMATICS/ISOPERIMETRIC AND ISODIAMETRIC FUNCTIONS OF GROUPS/NO FIELD/KEYWORDS PLUS: WORD PROBLEM; INEQUALITIES

49/BIRGET JC/2002/ANNALS OF MATHEMATICS/ISOPERIMETRIC FUNCTIONS OF GROUPS AND COMPUTATIONAL- COMPLEXITY OF THE WORD PROBLEM/NO FIELD/KEYWORDS PLUS: INEQUALITIES

425/GRIMALDI R/2003/JOURNAL DE MATHEMATIQUES PURES ET APPLIQUEES/FILLING AND SURFACES OF REVOLUTION/NO FIELD/KEYWORDS PLUS: ISOPERIMETRIC INEQUALITY; CURVATURE

CLUSTER 30

33/KOZLOVSKI OS/2003/ANNALS OF MATHEMATICS/AXIOM-A MAPS ARE DENSE IN THE SPACE OF UNIMODAL MAPS IN THE C-K TOPOLOGY/NO FIELD/KEYWORDS PLUS: DYNAMICS; POLYNOMIALS; MAPPINGS; SET

58/LYUBICH M/2002/ANNALS OF MATHEMATICS/ALMOST EVERY REAL QUADRATIC MAP IS EITHER REGULAR OR STOCHASTIC/NO FIELD/KEYWORDS PLUS: NON-LINEAR TRANSFORMATIONS; ONE-DIMENSIONAL MAPS; S-UNIMODAL MAPS; DYNAMICS; POLYNOMIALS; RENORMALIZATION; UNIVERSALITY; ITERATIONS; ATTRACTORS; FAMILIES

182/AVILA A/2003/INVENTIONES MATHEMATICAE/REGULAR OR STOCHASTIC DYNAMICS IN REAL ANALYTIC FAMILIES OF UNIMODAL MAPS/NO FIELD/KEYWORDS PLUS: ONE-DIMENSIONAL DYNAMICS; QUADRATIC POLYNOMIALS; HOLOMORPHIC MOTIONS; INVARIANT-MEASURES; HYPERBOLICITY; INTERVAL; BOUNDS

CLUSTER 31

274/TERASOMA T/2002/INVENTIONES MATHEMATICAE/MIXED TATE MOTIVES AND MULTIPLE ZETA VALUES/NO FIELD/NO FIELD

298/ELBAZVINCENT P/2002/INVENTIONES MATHEMATICAE/MILNOR K-THEORY OF RINGS, HIGHER CHOW GROUPS AND APPLICATIONS/NO FIELD/KEYWORDS PLUS: ALGEBRAIC CYCLES; HOMOLOGY

790/KAHN B/2002/MATHEMATISCHE ANNALEN/THE GEISSER-LEVINE METHOD REVISITED AND ALGEBRAIC CYCLES OVER A FINITE-FIELD/NO FIELD/KEYWORDS PLUS: BLOCH-KATO CONJECTURE; MILNOR K-THEORY; CHARACTERISTIC-P; COHOMOLOGY

CLUSTER 32

117/LUO WZ/2003/COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS/MASS EQUIDISTRIBUTION FOR HECKE EIGENFORMS/NO FIELD/KEYWORDS PLUS: AUTOMORPHIC L-FUNCTIONS; COEFFICIENTS; FORMS

267/DUKE W/2002/INVENTIONES MATHEMATICAE/THE SUBCONVEXITY PROBLEM FOR ARTIN L-FUNCTIONS/NO FIELD/KEYWORDS PLUS: AUTOMORPHIC L-FUNCTIONS; FOURIER COEFFICIENTS; MODULAR-FORMS; WEIGHT; BOUNDS; SUMS

725/HARCOS G/2003/MATHEMATISCHE ANNALEN/AN ADDITIVE PROBLEM IN THE FOURIER COEFFICIENTS OF CUSP FORMS/NO FIELD/KEYWORDS PLUS: HALF-INTEGRAL WEIGHT; SELBERG L-FUNCTIONS; DIVISOR PROBLEM; MODULAR-FORMS; SUMS; REPRESENTATIONS; OPERATORS

CLUSTER 33

304/BRIDGELAND T/2002/INVENTIONES MATHEMATICAE/FLOPS AND DERIVED CATEGORIES/NO FIELD/NO FIELD

507/NAMIKAWA Y/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/MUKAI FLOPS AND DERIVED CATEGORIES/NO FIELD/NO FIELD

647/CALDARARU A/2002/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/DERIVED CATEGORIES OF TWISTED SHEAVES ON ELLIPTIC THREEFOLDS/NO FIELD/KEYWORDS PLUS: FOURIER-MUKAI TRANSFORMS

CLUSTER 34

314/BERGER L/2002/INVENTIONES MATHEMATICAE/P-ADIC REPRESENTATION AND DIFFERENTIAL-EQUATIONS/NO FIELD/KEYWORDS PLUS: CRYSTALLINE REPRESENTATIONS; IWASAWA THEORY; INDEX THEOREM; F-ISOCRYSTALS; LOCAL-FIELD; COHOMOLOGY; EXTENSIONS; CURVE

315/ANDRE Y/2002/INVENTIONES MATHEMATICAE/HASSE-ARF FILTRATIONS AND P-ADIC MONODROMY/NO FIELD/KEYWORDS PLUS: F-ISOCRYSTALS; DIFFERENTIAL-EQUATIONS; INDEX THEOREM; GALOIS THEORY; REPRESENTATIONS; CURVE

316/MEBKHOUT Z/2002/INVENTIONES MATHEMATICAE/P-ADIC ANALOG OF THE TURRITTIN THEOREM AND THE THEOREM OF P- ADIC MONODROMY/NO FIELD/KEYWORDS PLUS: DIFFERENTIAL-EQUATIONS; INDEX THEOREM; F- ISOCRYSTALS; REPRESENTATIONS; COHOMOLOGY; OPERATORS; MODULES

CLUSTER 35

90/MARTEL Y/2002/ANNALS OF MATHEMATICS/STABILITY OF BLOW-UP PROFILE AND LOWER BOUNDS FOR BLOW-UP RATE FOR THE CRITICAL GENERALIZED KDV EQUATION/NO FIELD/KEYWORDS PLUS: DE-VRIES EQUATION; KORTEWEG-DEVRIES EQUATION

333/COLLIANDER J/2003/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/SHARP GLOBAL WELL-POSEDNESS FOR KDV AND MODIFIED KDV ON R AND T/AUTHOR KEYWORDS: KORTEWEG-DE VRIES EQUATION; NONLINEAR DISPERSIVE EQUATIONS; BILINEAR ESTIMATES; MULTILINEAR HARMONIC ANALYSIS/KEYWORDS PLUS: KORTEWEG-DEVRIES EQUATION; SEMILINEAR WAVE- EQUATIONS; ILL-POSEDNESS; DISPERSIVE EQUATIONS; EXISTENCE; SYSTEMS; TIME; L-2

361/MARTEL Y/2002/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/BLOW-UP IN FINITE-TIME AND DYNAMICS OF BLOW-UP SOLUTIONS FOR THE L-2-CRITICAL GENERALIZED KDV EQUATION/AUTHOR KEYWORDS: CRITICAL KDV EQUATION; FINITE TIME BLOW UP; BLOW UP RATE/KEYWORDS PLUS: KORTEWEG-DEVRIES EQUATION; DE-VRIES EQUATION

CLUSTER 36

416/MALCHIODI A/2002/JOURNAL DE MATHEMATIQUES PURES ET APPLIQUEES/A PERTURBATION RESULT FOR THE WEBSTER SCALAR CURVATURE PROBLEM ON THE CR SPHERE/AUTHOR KEYWORDS: WEBSTER CURVATURE; PERTURBATION METHODS; SUBELLIPTIC EQUATIONS/KEYWORDS PLUS: HEISENBERG-GROUP; SEMILINEAR EQUATIONS; VARIATIONAL APPROACH; YAMABE PROBLEM; HOMOCLINICS; LAPLACIAN; EXISTENCE; COMPLEX

759/DANCER EN/2003/MATHEMATISCHE ANNALEN/REAL ANALYTICITY AND NON-DEGENERACY/NO FIELD/KEYWORDS PLUS: POSITIVE SOLUTIONS; NONLINEAR EQUATIONS; ELLIPTIC-EQUATIONS; SCALAR CURVATURE; DOMAIN SHAPE; UNIQUENESS; NUMBER; BIFURCATION; WAVES

845/AMBROSETTI A/2002/MATHEMATISCHE ANNALEN/ON THE YAMABE PROBLEM AND THE SCALAR CURVATURE PROBLEMS UNDER BOUNDARY-CONDITIONS/NO FIELD/KEYWORDS PLUS: S-N; MEAN-CURVATURE; PERTURBATION; MANIFOLDS

CLUSTER 37

479/NAKAMAYE M/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/SESHADRI CONSTANTS AND THE GEOMETRY OF SURFACES/NO FIELD/NO FIELD

756/HWANG JM/2003/MATHEMATISCHE ANNALEN/SESHADRI-EXCEPTIONAL FOLIATIONS/NO FIELD/KEYWORDS PLUS: CONSTANTS; SURFACES

808/OGUISO K/2002/MATHEMATISCHE ANNALEN/SESHADRI CONSTANTS IN A FAMILY OF SURFACES/NO FIELD/KEYWORDS PLUS: VARIETIES

CLUSTER 38

186/WLODARCZYK J/2003/INVENTIONES MATHEMATICAE/TOROIDAL VARIETIES AND THE WEAK FACTORIZATION THEOREM/NO FIELD/KEYWORDS PLUS: GEOMETRIC INVARIANT-THEORY; BIRATIONAL MAPS; BLOWING-UP; SINGULARITIES; SURFACES; CHARACTERISTIC-0; THREEFOLDS; MORPHISMS; BUNDLES; RINGS

337/KAWAKITA M/2003/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/GENERAL ELEPHANTS OF 3-FOLD DIVISORIAL CONTRACTIONS/AUTHOR KEYWORDS: GENERAL ELEPHANT; DIVISORIAL CONTRACTION/KEYWORDS PLUS: 3-DIMENSIONAL TERMINAL SINGULARITIES

358/ABRAMOVICH D/2002/JOURNAL OF THE AMERICAN MATHEMATICAL SOCIETY/TORIFICATION AND FACTORIZATION OF BIRATIONAL MAPS/NO FIELD/KEYWORDS PLUS: GEOMETRIC INVARIANT-THEORY; BLOWING-UP; SINGULARITIES; SURFACES; CHARACTERISTIC-0; RESOLUTION; THREEFOLDS; MORPHISMS; VARIETIES; BUNDLES

CLUSTER 39

307/ETINGOF P/2002/INVENTIONES MATHEMATICAE/SYMPLECTIC REFLECTION ALGEBRAS, CALOGERO-MOSER SPACE, AND DEFORMED HARISH-CHANDRA HOMOMORPHISM/NO FIELD/KEYWORDS PLUS: QUANTUM INTEGRABLE SYSTEMS; AFFINE HECKE ALGEBRAS; KAC-MOODY ALGEBRAS; DUALIZING COMPLEXES; QUIVER VARIETIES; WEYL ALGEBRA; LIE-ALGEBRAS; OPERATORS; THEOREM; RINGS

672/NAKAJIMA H/2003/MATHEMATISCHE ANNALEN/REFLECTION FUNCTORS FOR QUIVER VARIETIES AND WEYL GROUP- ACTIONS/NO FIELD/KEYWORDS PLUS: ALE GRAVITATIONAL INSTANTONS; HYPER-KAHLER QUOTIENTS; KAC-MOODY ALGEBRAS; SPACES; SINGULARITIES; CONNECTIONS; MODULI

764/CRAWLEYBOEVEY W/2003/MATHEMATISCHE ANNALEN/NORMALITY OF MARSDEN-WEINSTEIN REDUCTIONS FOR REPRESENTATIONS OF QUIVERS/NO FIELD/KEYWORDS PLUS: KLEINIAN SINGULARITIES; ALE SPACES; ALGEBRAS; DEFORMATIONS; GEOMETRY

CLUSTER 40

66/DEBACKER S/2002/ANNALS OF MATHEMATICS/PARAMETRIZING NILPOTENT ORBITS VIA BRUHAT-TITS THEORY/NO FIELD/KEYWORDS PLUS: MINIMAL K-TYPES; P-ADIC GROUPS

251/MOEGLIN C/2003/INVENTIONES MATHEMATICAE/STABLE PACKS OF TEMPERED REPRESENTATIONS AND OF UNIPOTENT REDUCTION FOR SO(2N+1)/NO FIELD/KEYWORDS PLUS: CHARACTER SHEAVES; COHOMOLOGY

556/MURNAGHAN F/2003/JOURNAL FUR DIE REINE UND ANGEWANDTE MATHEMATIK/LOCAL CHARACTER EXPANSIONS OF ADMISSIBLE REPRESENTATIONS OF P-ADIC GENERAL LINEAR-GROUPS/NO FIELD/KEYWORDS PLUS: MINIMAL K-TYPES; FOURIER-TRANSFORM; SHALIKA GERMS; FIELD; GL(N); GLN

APPENDIX 8

THE COMPARISON OF TWO PARTITIONS IN CASE 3

The Dispersion of Articles over Clusters for Two Partitions.

In the following table, columns A-D show the dispersion of articles in clusters generated by the field expert over the clusters generated by the complete link cluster method whereas columns E-H show the dispersion of articles in clusters generated by the complete link cluster method over the clusters generated by the field expert. Cluster number "0" indicates articles not clustered by the field expert on grounds of insufficent information in the bibliographic descriptions.

A	B	C	D	E	F	G	H
Complete Doc. nr.	Complete Clu.nr.	Expert Doc.nr.	Expert Clu.nr.	Complete Doc. nr.	Complete Clu.nr.	Expert Doc.nr.	Expert Clu.nr.
274	31	274	0	475	1	475	2
337	38	337	0	711	1	711	34
628	13	628	0	712	1	712	34
779	19	779	1	205	2	205	12
475	1	475	2	207	2	207	12
267	32	267	2	463	2	463	16
230	27	230	3	454	3	454	10
318	27	318	3	686	3	686	36
298	31	298	3	788	3	788	4
788	3	788	4	215	4	215	16
186	38	186	5	559	4	559	15
358	38	358	5	665	4	665	30
859	12	859	6	803	4	803	15
523	19	523	6	240	5	240	18
716	19	716	6	279	5	279	15
693	23	693	6	656	5	656	15
830	23	830	6	61	6	61	31
549	25	549	6	192	6	192	31
94	21	94	7	193	6	193	19
312	17	312	8	173	7	173	25
472	17	472	8	798	7	798	25
309	23	309	9	857	7	857	25
454	3	454	10	220	8	220	19
261	16	261	10	281	8	281	19
867	16	867	10	499	8	499	19
200	12	200	11	437	9	437	26
8	22	8	11	630	9	630	26
205	2	205	12	738	9	738	26
207	2	207	12	99	10	99	26
58	30	58	12	494	10	494	24
182	30	182	12	535	10	535	24
869	13	869	13	269	11	269	24
72	18	72	13	295	11	295	24
304	33	304	13	334	11	334	24
507	33	507	13	84	12	84	16
713	25	713	14	200	12	200	11

559	4	559	15
803	4	803	15
279	5	279	15
656	5	656	15
242	13	242	15
571	28	571	15
790	31	790	15
463	2	463	16
215	4	215	16
84	12	84	16
804	15	804	16
871	15	871	16
319	22	319	16
725	32	725	16
212	26	212	17
243	26	243	17
321	26	321	17
338	26	338	17
240	5	240	18
79	27	79	18
11	28	11	18
193	6	193	19
220	8	220	19
281	8	281	19
499	8	499	19
363	14	363	19
667	14	667	19
677	14	677	19
18	18	18	20
73	18	73	20
349	18	349	20
364	18	364	20
315	34	315	20
316	34	316	20
556	40	556	20
264	28	264	21
333	35	333	22
361	35	361	22
307	39	307	23
672	39	672	23
494	10	494	24
535	10	535	24
269	11	269	24
295	11	295	24
334	11	334	24
486	16	486	24
564	16	564	24
12	17	12	24
15	24	15	24
108	24	108	24
787	24	787	24
859	12	859	6
242	13	242	15
628	13	628	0
869	13	869	13
363	14	363	19
667	14	667	19
677	14	677	19
635	15	635	27
804	15	804	16
871	15	871	16
261	16	261	10
486	16	486	24
564	16	564	24
867	16	867	10
12	17	12	24
312	17	312	8
472	17	472	8
18	18	18	20
21	18	21	31
72	18	72	13
73	18	73	20
349	18	349	20
364	18	364	20
523	19	523	6
642	19	642	33
716	19	716	6
779	19	779	1
276	20	276	29
426	20	426	32
875	20	875	29
94	21	94	7
124	21	124	27
150	21	150	27
162	21	162	27
379	21	379	32
8	22	8	11
319	22	319	16
580	22	580	28
309	23	309	9
693	23	693	6
830	23	830	6
15	24	15	24
69	24	69	26
108	24	108	24
787	24	787	24
154	25	154	31
549	25	549	6
713	25	713	14
212	26	212	17
243	26	243	17
321	26	321	17

416	36	416	24
845	36	845	24
479	37	479	24
756	37	756	24
808	37	808	24
90	35	90	25
173	7	173	25
798	7	798	25
857	7	857	25
437	9	437	26
630	9	630	26
738	9	738	26
99	10	99	26
69	24	69	26
759	36	759	26
635	15	635	27
124	21	124	27
150	21	150	27
162	21	162	27
48	29	48	27
49	29	49	27
425	29	425	27
580	22	580	28
276	20	276	29
875	20	875	29
117	32	117	29
665	4	665	30
61	6	61	31
192	6	192	31
21	18	21	31
154	25	154	31
314	34	314	31
764	39	764	31
251	40	251	31
426	20	426	32
379	21	379	32
33	30	33	32
642	19	642	33
647	33	647	33
711	1	711	34
712	1	712	34
66	40	66	35
686	3	686	36
338	26	338	17
79	27	79	18
230	27	230	3
318	27	318	3
11	28	11	18
264	28	264	21
571	28	571	15
48	29	48	27
49	29	49	27
425	29	425	27
33	30	33	32
58	30	58	12
182	30	182	12
274	31	274	0
298	31	298	3
790	31	790	15
117	32	117	29
267	32	267	2
725	32	725	16
304	33	304	13
507	33	507	13
647	33	647	33
314	34	314	31
315	34	315	20
316	34	316	20
90	35	90	25
333	35	333	22
361	35	361	22
416	36	416	24
759	36	759	26
845	36	845	24
479	37	479	24
756	37	756	24
808	37	808	24
186	38	186	5
337	38	337	0
358	38	358	5
307	39	307	23
672	39	672	23
764	39	764	31
66	40	66	35
251	40	251	31
556	40	556	20

PUBLIKATIONER UR SERIEN SKRIFTER FRÅN VALFRID

Enmark, Romulo: Defining the Library's Activities
(ISBN 91-971457-1-X) International Publications: 1

Biblioteksstudier. Folkbibliotek i flervetenskaplig belysning. Red. Romulo Enmark.
(ISBN 91-971457-0-X) Skriftserien ; 1

Biblioteken och framtiden, del I. Red. Romulo Enmark.
(ISBN 91-971457-1-8) Skriftserien ; 2

Biblioteken och framtiden, del II. Red. Lars Seldén.
(ISBN 91-971457-2-6) Skriftserien ; 3

Hjørland, Birger: Emnerepræsentation og informationssøgning.
(ISBN 91-971457-3-4) Slut

Hjørland, Birger: Emnerepræsentation og informationssøgning. 2. uppl. med register.
(ISBN 91-971457-4-2) Skriftserien ; 4

Biblioteken, kulturen och den sociala intelligensen. Red. Lars Höglund.
(ISBN 91-971457-5-0) Skriftserien ; 5

Hjørland, Birger: Faglitteratur. Kvalitet, vurdering og selektion.
(ISBN 91-971457-6-9) Skriftserien ; 6

Limberg, Louise: Skolbiblioteksmodeller. Utvärdering av ett utvecklingsprojekt i Örebro län. (ISBN 91-971457-7-7) Slut

Hjørland, Birger: Faglitteratur. Kvalitet, vurdering og selektion. 2. rev.udgave.
(ISBN 91-971457-8-5) Skriftserien ; 8

Pettersson, Rune: Verbo-visual Communication – Presentation of Clear Messages for Information and Learning.
(ISBN 91-971457-9-3) Skriftserien ; 9

Pettersson, Rune: Verbo-visual Communication – 12 Selected Papers
(ISBN 91-973090-0-1) Skriftserien ; 10

Limberg, Louise: Skolbiblioteksmodeller. Utvärdering av ett utvecklingsprojekt i Örebro län.
(ISBN 91-973090-1-X) Skriftserien ; 11 (nytryck av nr 7, bilaga inne i boken)

Barnbibliotek och informationsteknik. Elektroniska medier för barn och ungdomar på folkbibliotek. Red. Anette Eliasson, Staffan Lööf, Kerstin Rydsjö.
(ISBN 91-973090-2-8) Skriftserien ; 12

Folkbildning och bibliotek? På spaning efter spår av folkbildning och livslångt lärande i biblioteksvärlden. Red. Maj Klasson.
(ISBN 91-973090-3-6) Skriftserien ; 13

Zetterlund, Angela: Utvärdering och folkbibliotek : En studie av utvärderingens teori och praktik med exempel från folkbibliotekens förändrings- och utvecklingsprojekt
(ISBN 91-973090-4-4) Slut

Myrstener, Mats: På väg mot ett stadsbibliotek. Folkbiblioteksväsendets framväxt i Stockholm t o m 1927.
(ISBN 91-973090-5-2) Skriftserien ; 15

Limberg, Louise: Att söka information för att lära. En studie av samspel mellan informationssökning och lärande
(ISBN 91-973090-6-0, Slut)
(ISBN 91-89416-04-X, nytryck 2001 och 2003) Skriftserien ; 16

Hansson, Joacim: Om folkbibliotekens ideologiska identitet. En diskursstudie
(ISBN 91-973090-7-9) Slut

Gram, Magdalena: Konstbiblioteket : en krönika och en fallstudie
(ISBN 91-973090-8-7) Skriftserien ; 18

Hansson, Joacim: Klassifikation, bibliotek och samhälle. En kritisk hermeneutisk studie av "Klassifikationssystem för svenska bibliotek"
(ISBN 91-973090-9-5) Skriftserien ; 19

Seldén, Lars: Kapital och karriär. Informationssökning i forskningens vardagspraktik.
(ISBN 91-89416-00-7 Slut)
(ISBN 91-89416-08-2, nytryck 2004) Skriftserien ; 20

Edström, Göte: Filter, raster, mönster. Litteraturguide i teori- och metodlitteratur för biblioteks- och informationsvetenskap och angränsande ämnen inom humaniora och samhällsvetenskap.
(ISBN 91-89416-01-5) Slut

Röster. Biblioteksbranden i Linköping. Red. Maj Klasson
(ISBN 91-89416-02-3) Skriftserien ; 22

Stenberg, Catharina: Litteraturpolitik och bibliotek. En kulturpolitisk analys av bibliotekens litteraturförvärv speglad i Litteraturutredningen L 68 och Folkbiblioteksutredningen FB 80.
(ISBN 91-89416-03-1) Skriftserien ; 23

Edström, Göte: Filter, raster, mönster. Litteraturguide i teori- och metodlitteratur för biblioteks- och informationsvetenskap och angränsande ämnen inom humaniora och samhällsvetenskap. Andra aktualiserade och utökade upplagan.
(ISBN 91-89416-05-8) Skriftserien ; 24

Sundin, Olof: Informationsstrategier och yrkesidentiteter - en studie av sjuksköterskors relation till fackinformation vid arbetsplatsen.
(ISBN 91-89416-06-6) Skriftserien ; 25

Hessler, Gunnel: Identitet och förändring - en studie av ett universitetsbibliotek och dess självproduktion.
(ISBN 91-89416-07-4) Skriftserien ; 26

Zetterlund, Angela: Att utvärdera i praktiken - en retrospektiv fallstudie av tre program för lokal folkbiblioteksutveckling.
(ISBN 91-89416-09-0) Skriftserien ; 27

Ahlgren, Per: The Effects on Indexing Strategy-Query Term Combination on Retrieval Effectiveness in a Swedish Full Text Database.
(ISBN 91-89416-10-4) Skriftserien ; 28

Thórsteinsdóttir, Gudrun: The Information Seeking Behaviour of Distance Students. A Study of Twenty Swedish Library and Information Science Students
(ISBN 91-89416-11-2) Skriftserien ; 29

BO JARNEVING

THE COMBINED APPLICATION OF BIBLIOGRAPHIC COUPLING AND THE COMPLETE LINK CLUSTER METHOD IN BIBLIOMETRIC SCIENCE MAPPING

Bibliometrics is the quantitative study of patterns derived from the production and use of publications where mathematical and statistical methods are applied. The focus of this thesis connects to previous research in bibliometric science mapping and citation indexing. A citation index facilitates the retrieval of documents associated through citation links whereas the objective of citation based science mapping is to reveal the cognitive structures of science and to provide scientists with information.

In the empirical study presented, a method for the partition of document populations, the complete link cluster method and a method for the measuring of similarity between research articles, bibliographic coupling, were combined to a method preliminarily fit for science mapping purposes. The aim of the study was to evaluate this method and to find its area of application.

Findings showed that the proposed method has the capability to identify and map current and coherent research themes on the level of a single research field as well as in a multidisciplinary environment. However, based on theoretical considerations as well as on empirical findings, it was concluded that it would not suffice as a standard science mapping method where exhaustive depictions of specialties' cognitive structures are aimed at. On these grounds, it was concluded that the area of application of the proposed method should be scientific information provision and that it would be complementary to traditional citation indexing.

Bo Jarneving is a member of the teaching staff at the Department of Library and Information Science/School of Library and Information Science, University College of Borås and Göteborg University. *The Combined Application of Bibliographic Coupling and the Complete Link Cluster Method in Bibliometric Science Mapping* is his Doctoral Thesis.

Department of Library and Information Science/
Swedish School of Library and Information Science
University College of Borås and Göteborg University

ISBN 91-89416-12-0
ISSN 1103-6990