From the Department of Learning, Informatics, Management and Ethics
Karolinska Institutet, Stockholm, Sweden

Mapping the structure of science through clustering in citation networks: granularity, labeling and visualization

Peter Sjögårde

Karolinska Institutet

Stockholm 2023

Published by Karolinska Institutet.
Printed by Universitetsservice US-AB, 2023.

ISBN 978-91-8017-025-3
Cover illustration: Map of science based on about 20 million biomedical research publications from PubMed. Colors show research fields that are more (yellow) or less (blue) clinically oriented.

Mapping the structure of science through clustering in citation networks: granularity, labeling and visualization
THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Peter Sjögårde

Principal Supervisor:
Sabine Koch
Karolinska Institutet
Department of Learning, Informatics, Management and Ethics
Health Informatics Centre

Co-supervisors:
Carl Johan Sundberg
Karolinska Institutet
Department of Learning, Informatics, Management and Ethics
Medical Management Centre

Per Ahlgren
Uppsala University
Department of Statistics

Ludo Waltman
Leiden University
Centre for Science and Technology Studies

Opponent:
Martin Rosvall
Umeå University
Department of Physics

Examination Board:
Jussi Karlgren
Helsinki University
Language Technology, Faculty of Arts

Koraljka Golub
Linnaeus University
Department of Cultural Sciences

Kim Holmberg
University of Turku
Department of Information Studies

POPULAR SCIENCE SUMMARY

Imagine that you are asked to organize and describe the landscape of all biomedical research published during the last 30 years – a set of about 20 million publications. You are also told to provide an overview showing the main research fields and the possibility to zoom into more narrow areas to explore research of interest in more detail. This is the task I focus on in this thesis.

However, I did not start from scratch. Like Isaac Newton, I have been "standing on the shoulders of giants", meaning that I have built my work on previous research, in which methods to group publications based on how they are related have been elaborated.

I will briefly explain how this is done. In research publications, researchers use references to relate their work to the works of others, as well as to their own former works. Such references can be used to create links, or "citations", between publications. Citation relations can be used to create networks of publications. The mentioned set of about 20 million biomedical research publications includes about half a billion citation relations. Algorithms have been elaborated that can separate groups of publications that are densely interconnected from other areas of the network. This makes it possible to divide the network of publications into smaller groups, or "clusters". A hierarchical classification is created when applying this method at different levels, including broad and coarse levels as well as narrow and highly granular levels.

So, what has been my contribution?

1. First of all, I have paid attention to the fundamentals of this approach. What kind of groups are created using such a method? What do we mean by terms such as "research fields", "research topics" or "research specialties" and how do clusters correspond to such terms? In short, my answer is that research fields are areas of research that researchers focus on. Clusters represent focus areas of research, because researchers refer to work which is generally within the same focus area.
2. Secondly, I have focused on how to adjust the size of clusters at different levels. The clustering algorithm includes a parameter which determines the size of clusters. In the first two articles of the thesis, I propose a method to adjust this parameter to create two levels: (1) the level of "research topics", which includes narrow and detailed focus areas addressed in research publications (up to a few hundred thousand areas), and (2) the level of "research specialties", which includes broader areas of research addressed by research communities (a few thousand areas).
3. Thirdly, I have studied how clusters can be labeled. Without labels, it is almost impossible for a user to understand the contents of a cluster, in particular if the cluster is large. In the third article, I propose which data to use for labeling of clusters of different size, using broad terms such as "psychiatry" or "orthopedics" for large

clusters and specific terms such as “bipolar disorder” or “metal joint prosthesis” for narrow clusters.

4. Lastly, I have prosed a way to present the classification visually. This method uses large clusters to provide an overview of the biomedical research landscape and small clusters to make it possible for users to zoom into research fields and get details about the topics addressed within each focus area.

In summary, I have improved citation-based classifications by providing a logic behind the different levels of the classification, labeling clusters at different levels, and by making it possible to visually navigate the classification. These improvements make the classification more useful. The obtained classification can be used to study different aspects of the biomedical research landscape, for example to find out in which areas artificial intelligence is applied[1] or to determine which areas are more clinically oriented[2]. It can also be used to identify research from a particular researcher, journal, or organization, and to explore how this research fits into the overall biomedical research landscape. I have, for example, created maps of the publications from Karolinska institutet[3] and of Covid-19 related research[4]. A classification of biomedical research literature[5] has also been published online so it can be used by others.

In this thesis, I have demonstrated how the biomedical research landscape can be organized and described in a way that provides both overview and detail.

[1] https://petersjogarde.github.io/papers/ai/world/index.html
[2] https://petersjogarde.github.io/pm_classification/2023feb/basemap/clinical/index.html
[3] https://petersjogarde.github.io/papers/hiervis/sthlm_trio/ki/index.html
[4] https://petersjogarde.github.io/papers/hiervis/covid_v2/pubs/index.html
[5] https://figshare.com/collections/PubMed_Classification/5610971

ABSTRACT

The science system is large, and millions of research publications are published each year. Within the field of scientometrics, the features and characteristics of this system are studied using quantitative methods. Research publications constitute a rich source of information about the science system and a means to model and study science on a large scale. The classification of research publications into fields is essential to answer many questions about the features and characteristics of the science system.

Comprehensive, hierarchical, and detailed classifications of large sets of research publications are not easy to obtain. A solution for this problem is to use network-based approaches to cluster research publications based on their citation relations. Clustering approaches have been applied to large sets of publications at the level of individual articles (in contrast to the journal level) for about a decade. Such approaches are addressed in this thesis. I call the resulting classifications "algorithmically constructed, publications-level classifications of research publications" (ACPLCs).

The aim of the thesis is to improve interpretability and utility of ACPLCs. I focus on some issues that hitherto have not received much attention in the previous literature: (1) *Conceptual framework*. Such a framework is elaborated throughout the thesis. Using the social science citation theory, I argue that citations contextualize and position publications in the science system. Citations may therefore be used to identify research fields, defined as focus areas of research at various granularity levels. (2) *Granularity levels corresponding to conceptual framework*. In Articles I and II, a method is proposed on how to adjust the granularity of ACPLCs in order to obtain clusters corresponding to research fields at two granularity levels: topics and specialties. (3) *Cluster labeling*. Article III addresses labeling of clusters at different semantic levels, from broad and large to narrow and small, and compares the use of data from various bibliographic fields and different term weighting approaches. (4) *Visualization*. The methods resulting from Articles I-III are applied in Article IV to obtain a classification of about 19 million biomedical articles. I propose a visualization methodology that provides overview of the classification, using clusters at coarse levels, as well as the possibility to zoom into details, using clusters at a granular level.

In conclusion, I have improved interpretability and utility of ACPLCs by providing a conceptual framework, adjusting granularity of clusters, labeling clusters and, finally, by visualizing an ACPLC in a way that provides both overview and detail. I have demonstrated how these methods can be applied to obtain ACPLCs that are useful to, for example, identify and explore focus areas of research.

LIST OF SCIENTIFIC PAPERS

I. Sjögårde, P., & Ahlgren, P. (2018). Granularity of algorithmically constructed publication-level classifications of research publications: Identification of topics. *Journal of Informetrics*, *12*(1), 133–152. https://doi.org/10.1016/j.joi.2017.12.006

II. Sjögårde, P., & Ahlgren, P. (2020). Granularity of algorithmically constructed publication-level classifications of research publications: Identification of specialties. *Quantitative Science Studies*, *1*(1), 207–238. https://doi.org/10.1162/qss_a_00004

III. Sjögårde, P., Ahlgren, P., & Waltman, L. (2021). Algorithmic labeling in hierarchical classifications of publications: Evaluation of bibliographic fields and term weighting approaches. *Journal of the Association for Information Science and Technology*, *72*(7), 853–869. https://doi.org/10.1002/asi.24452

IV. Sjögårde, P. (2022). Improving overlay maps of science: Combining overview and detail. *Quantitative Science Studies*, *3*(4), 1097–1118. https://doi.org/10.1162/qss_a_00216

ADDITIONAL PAPERS

- Ahlgren, P., Colliander, C., & Sjögårde, P. (2018). Exploring the relation between referencing practices and citation impact: A large-scale study based on Web of Science data. *Journal of the Association for Information Science and Technology*, *69*(5), 728–743. https://doi.org/10.1002/asi.23986

- Sjögårde, P., & Didegah, F. (2022). The association between topic growth and citation impact of research publications. *Scientometrics*, *127*(4), 1903–1921. https://doi.org/10.1007/s11192-022-04293-x

- Sjögårde, P. (2022). The effect of the rapid growth of covid-19 publications on citation indicators. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *Proceedings of the 26th International Conference on Science and Technology Indicators*. https://doi.org/10.5281/zenodo.7142412

- Sjögårde, P. (2022). Exploring user needs in relation to algorithmically constructed classifications of publications: A case study. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *Proceedings of the 26th International Conference on Science and Technology Indicators*. https://doi.org/10.5281/zenodo.6959252

PUBLISHED DATASETS

- Sjögårde, Peter (2021): *PubMed Classification*. figshare. Collection. https://doi.org/10.6084/m9.figshare.c.5610971.v3

CONTENTS

LIST OF ABBREVIATIONS

ACPLC	Algorithmically Constructed Publication-Level Classification
ARI	Adjusted Rand Index
CPM	Constant Potts Model
JSD	Jensen Shannon Divergence
KI	Karolinska Institutet
KTH	KTH Royal Institute of Technology
MeSH	Medical Subject Headings
NIH OCC	National Institute of Health Open Citation Collection
RA	PubMed Related Article measure
SMJC	Science-Metrix Journal Classification
SSCT	Social Science Citation Theory
SU	Stockholm University
TF-IDF	Term Frequency-Inverse Document Frequency
TFS	Term Frequency to Specificity ratio
WvE	Waltman and van Eck term weighting approach

LIST OF DEFINITIONS

Bibliometrics	The quantitative study of publication collections.
Scientometrics	The quantitative study of the features and characteristics of science and scientific research.
Classification	The process of distinguishing and distributing kinds of 'things' into different groups as well as the resulting delineation of these 'things' into classes (Hjørland, 2020).
Classification schema	A list of empty classes and their relations used for, or resulting from, classification (in the verb sense of the term).
Unsupervised classification	The task to algorithmically perform classification without the use of a predefined classification schema.
Research field	A focus area of research at any granularity level. I consider research topics, specialities and disciplines as research fields at various granularity levels (from narrow to broad).
Research topic	A focus area of research that constitutes a thematic context to which researchers relate research questions in research studies.
Research specialty	A focus area of research addressed by "*a self-organized network of researchers who tend to study the same research topics, attend the same conferences, read and cite each other's research papers and publish in the same journals.*" (Morris & van der Veer Martens, 2008)
Research discipline	A broad focus area of research consisting of multiple specialties.
Research community	A self-organized network of researchers addressing focus areas of research.
Research Concept	A bearer of linguistic meaning related to an aspect of a research publication.
Publication	A document made available as printed or electronic media, such as a journal article, conference paper, report, or book.
Bibliographic record	The set of metadata describing a publication, including for example title, abstract, year and references.

Reference list	The structured list of publications provided in the end of research publications.
Reference	A textual instance in a reference list referring to a publication.
Citation	The link between two publications created by matching a reference to a publication.
Clustering	The task to perform unsupervised classification of items based on their relatedness in networks.
Community detection	I use community detection synonymously to clustering.
Quality function	A function used for the evaluation of a clustering solution.
Optimization algorithm	An algorithm that obtains a clustering solution by optimizing the value of a quality function.
Publication-publication similarity measure	The measure used to calculate the similarity between two publications.
Normalization approach (in network context)	The approach taken to normalize citation relations, primarily to reduce differences between research fields and age of publications.

1 INTRODUCTION - MODELLING AND STUDYING THE STRUCTURE OF SCIENCE THROUGH PUBLICATIONS

In the science system, researchers organize themselves to address problems that are formulated within research communities, often responding to needs or discourses in the surrounding society (Morris & van der Veer Martens, 2008). Science has been evolving for centuries and has expanded its coverage and at the same time become more specialized and detailed as new fields of science have emerged, typically breaking out from existent fields. New research questions have been formulated and theories have triggered new approaches to address different issues. Novel methodologies have led to rapid and disruptive development and shifts in paradigms (Kuhn, 1996). The science system is dynamic and in continuing change and development (Sugimoto & Weingart, 2015). Is it possible to obtain an overview of such a complex system and model its structure and properties?

No researcher covers all areas of science. On the contrary, science is performed and communicated within communities having different focus areas (Börner et al., 2005; Crane, 1972). The communities are overlapping, and the borders are vague (Chubin, 1976; Havemann et al., 2017). They form core and peripheral communication structures, having dense communication in the core of the communities and more sparse communication in the periphery (de Solla Price & Beaver, 1966; Wedell et al., 2022). Furthermore, the communities are characterized by having a more intensive internal communication than their communication with other parts of the science system (Boyack, 2017; de Solla Price & Beaver, 1966). In this thesis, such features are used to model the structure of the science system.

One of the most important formal ways of communication in the science system is through publications, such as journal articles, conference papers, reports and books (Scharnhorst et al., 2012). In publications, findings are presented, theories are developed, and methods are described. Traces of collaboration are expressed by co-authorship, which is more likely to occur within a research community than between researchers of different communities (Newman, 2004a; H. White & McCain, 1998). The terminology in publications expresses focus areas and the subject orientation of communities, which is to some extent specific to each community (Callon et al., 1983). Furthermore, and importantly for this thesis, researchers situate themselves in their research context and their research community by referencing to other publications (de Solla Price, 1965b; Tahamtan & Bornmann, 2022). This practice relates publications to their historical foundation (Garfield, 1963), and their theoretical, methodological, and content-oriented context. Altogether research publications constitute a rich source of information about the science system and a means to simplify, model, and study the system at large scale (de Solla Price, 1965a; Garfield, 1963).

Bibliometrics is the quantitative study of publication collections and is essential in the field of scientometrics (or quantitative science studies). The delineation of publications into research fields is fundamental to a wide range of scientometric studies, for example to study co-

publishing in and between research fields, growth and emergence of research fields, differences in disciplinary orientation between organizations and countries, and shifts in topical orientation. A mosaic of manually obtained classifications or other categorizations intended to describe the subject orientation of publications exists. However, existing classifications and nomenclatures are insufficient for the study of a wide range of questions about the science system. Some classifications are too broad to provide detail, for example journal level classifications such as the Web of Science journal classification. Nomenclatures such as the Medical Subject Headings (MeSH) are not always optimal for the purpose of studying the science system because they were originally constructed for the purpose of searching or browsing publication collections and do not capture focus areas of research very well. It may be possible to obtain detailed classifications manually. However, such an approach is costly to maintain. To this point in time, classifications that are comprehensive, granular at its lowest level, easy to maintain and fit for the purpose of studying the structure of the science system are still missing.

In this thesis, I focus on algorithmic approaches to obtain classifications of research publications, in particular of journal articles in biomedicine. I concentrate on network-based approaches to obtain classifications, where the entities, or "nodes", of the networks in question are publications and each link, or "edge", between two nodes represents a citation from one publication to another publication, a so-called "direct citation" relation. A network can be divided into smaller groups using clustering (also called community-detection) algorithms. Different methods to delineate the nodes of a network into groups, or "clusters", exist. Their overall goal is to obtain clusters having a higher concentration of edges within the cluster than outside the clusters (Fortunato, 2010). In the case of citation networks, clusters represent dense areas of formal research communication. Such areas are likely to represent focus areas or research, where the publications of each cluster are likely to share some properties, for example problem definitions, ontologies, vocabulary, epistemology, methodology, geographical focus, or entities of examination (e.g., species, substances, or artefacts). Clustering of citation networks is therefore a promising approach to map and study the science system. However, I emphasize that it is hardly possible to create one single classification to model science and to answer all different research questions. The classification must be fit for the purpose of its use.

Algorithmic methods to create classifications of research publications have been developed since the 1960s. In recent years, clustering algorithms have been used to obtain large-scale classifications of research publications of high granularity. Such classifications can be used to get an overview of the science system, study interactions in the system and delineate science into fields at various levels of aggregation. In this thesis, I summarize my work on improving methods to obtain such classifications. I have studied large networks of research publications, their relations through citations, and how these networks can be used to obtain classifications of publications.

The thesis is restricted to approaches using clustering in large citation networks. Clustering is a type of unsupervised approach in the sense that it does not make use of a pre-existing classification schema. Moreover, I focus on publication-level classifications, in contrast to journal-level classifications, which is necessary to obtain classifications of high granularity.

1.1 AIM OF THE THESIS

The aim of the thesis is to improve interpretability and utility of algorithmically constructed, publications-level classifications of research publications (ACPLCs). To fulfill the aim, I focus on some issues that have not received much attention in the previous literature:

1. *Conceptual framework* – Several terms have been used in the scientometric literature to denote the groups of publications obtained by clustering, for example "disciplines", "research fields", "research areas", "topics" and "specialties". There is no consensus about these terms. How they relate to the clusters obtained from the chosen approaches has not been thoroughly addressed in the literature. A conceptual framework has been elaborated throughout the project and is outlined in this thesis frame.
2. *Granularity levels corresponding to conceptual framework* – The granularity of classifications is regularly determined by the user arbitrarily, for instance by setting the number of clusters to be obtained or by the choice of a resolution parameter value given to a clustering software. In Articles I and II, we address the question on how the granularity of ACPLCs can be adjusted so that clusters correspond to topics and specialties as these notions have been defined in the conceptual framework.
3. *Cluster labeling* – Without labels, it is very time consuming to interpret the subject orientation of clusters of research publications, in particular in large clusters. This restricts utility. Nevertheless, only a few studies have focused on labeling of algorithmically obtained clusters of research publications (Koopman et al., 2017; Velden, Boyack, et al., 2017; Velden, Yan, et al., 2017; Waltman & van Eck, 2012). In Article III, we address how labels can be assigned to clusters at different sematic levels.
4. *Visualization* – Visualization of classifications is necessary in many use cases, for example to make it possible for a user to get an overview or to explore a research field of interest. To my best knowledge, no scientometric study has incorporated several levels in a classification in the same visualization. In Article IV, I address how ACPLCs can be visualized to provide both overview and detail of the science system. This fourth study synthesizes the previous work in the project by the use of the methods resulting from Articles I-III.

The thesis is structured as follows. In Section 2, I give a historical background of the research literature of the use of citation networks for the classification of research publications. The conceptual framework is explained in Section 3. Data and methods are described in Section 4. The articles of the thesis are summarized in Section 0, followed by discussion in Section 6. Conclusions are made in Section 7.

2 BACKGROUND – HISTORY OF NETWORK-BASED CLASSIFICATION OF SCIENCE

The classification of science has a long history. Glänzel and Schubert (2003a) suggest that the task to classify science "into a disciplinary structure is at least as old as science itself". This may be true; classification of science goes back at least to ancient Greece. For example, Aristotle classified science as "theoretical", "practical", or "productive" (Hjørland, 2020) and words such as "mathematics", "astronomy", "philosophy" and "biology" stem from ancient Greece. Traditional library classification (the arrangement of documents into categories) goes back at least to the Alexandria library in the 3rd century. Several of the classifications used in many academic libraries today have a history of 100 years or more. For example, the "Dewey Decimal Classification" was first published in 1876. The idea of faceted classification is almost 100 years old and stems from the "Colon Classification" developed by Ranganathan in the first half of the 20th century (Hjørland, 2020). All these classifications have in common that they are created in a subjective manner, based on how the creator perceived the delineation of knowledge into categories. Modern classification schemas of this kind are developed collaboratively by multiple persons and organizations, often following sophisticated processes and criteria. Nevertheless, such classification schemas are still developed based on subjectively perceived categories and have been sparsely evaluated in the scientometric literature.

Conventionally, classifications of research publications are obtained by two steps, the first being the creation of a classification schema and the second the assignment of publications to the classes of the classification schema. This procedure is performed by librarians at university libraries or national libraries and is part of cataloguing procedures. The classifications have generally been created to enhance the possibilities to browse publications within a subject of interest to the user, specifically when the publication is not known by the user beforehand. The use of classifications in scientometric studies is often rather different from browsing the library shelf. In scientometric studies, classifications are primarily used for statistical analyses. For example, scientometric studies require classifications to:

- delineate publications by research fields in analyses (e.g., Ahlgren et al., 2018; Haunschild et al., 2018; Milanez et al., 2016; Muñoz-Écija et al., 2019; Sjögårde & Didegah, 2022).
- measure interdisciplinarity (e.g., Abramo et al., 2018; e.g. Engerer, 2017; Gerlach et al., 2018; Katz & Hicks, 1995; Morillo et al., 2001; Porter et al., 2007; Raan, 2005a; Q. Wang & Ahlgren, 2018; Q. Wang & Schneider, 2019).
- detect emerging topics (e.g., H. Small et al., 2014; Q. Wang, 2018; Xu et al., 2020).
- normalize citation indicators to research fields (e.g., Bornmann et al., 2013; Leydesdorff et al., 2013; Raan, 2005b; Ruiz-Castillo & Waltman, 2015).

Scientometric studies often need classifications that are comprehensive (in terms of coverage of the classified publications) and of high granularity which makes detailed large-scale studies possible.

Web of Science, and its predecessor the Science Citation Index, has been the pioneer data source for scientometric analyses. It was proposed by Eugene Garfield in 1955 and developed in the 1950s and 60s (Garfield, 1955). Journal categories were created in conjunction with the development of the Journal Citation Reports, which contain the well-known journal impact factor (Garfield, 1972).[6] The Web of Science journal classification has been widely used for scientometric studies. However, the classification has some properties that limit its usefulness:

1. Journal-level classifications are inevitably coarse because of the size and wide scope of many journals. Hence, the categories are not useful for studies of more narrow scope. Detailed studies of the subject orientation of a unit of analysis cannot be performed using the Web of Science journal classification.
2. The categories are predefined by the creator. Thereby, the categories reflect the disciplinary structure of science as subjectively perceived by the creator. The accuracy of the Web of Science journal classification has been convincingly contested (Klavans & Boyack, 2017a; Rafols & Leydesdorff, 2009; Q. Wang & Waltman, 2016).
3. Pre-defined categories adapt slowly to new developments in science because the classification schema must be manually updated for new categories to be included. This makes such classifications less suitable for detection of emerging research fields.

These weaknesses are not specific to the Web of Science journal classification. They apply to all classifications of journals into predefined categories, for example the All Science Journal Classification in Scopus.[7]

The use of controlled vocabularies, e.g., Medical Subject Headings (MeSH), is an alternative approach for the study of the science system. However, such nomenclatures have been created for the purpose of search and are less suitable for some analytical purposes.[8] Some MeSH terms correspond rather well to a research field, while others do not. For example, most publications with the MeSH term "carcinoma, merkel cell" (a type of skin cancer) belong to the same research field, or at least they can be grouped using clustering in citation networks. However, the MeSH term "arm" is very unspecific and is scattered over many research fields.

Methods for large-scale algorithmic classification of research publications have been developed to better meet the needs of quantitative studies of science. Network-based approaches have been dominating. They are unsupervised in the sense that they do not depend on pre-existing classification schemas or training data. Since they are recreated, they

[6] "Web of Science Core Collection Help". URL: https://images.webofknowledge.com/WOKRS56B5/help/WOS/hp_subject_category_terms_tasca.html [2020-08-24]
[7] "What is the complete list of Scopus Subject Areas and All Science Journal Classification Codes (ASJC)?". URL: https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/ [2020-08-24]
[8] It should be noted that MeSH can be used to calculate publication-publication similarity and for clustering. I here refer to the direct use of MeSH as a classification.

have the potential to capture changes in the science system. Furthermore, network-based approaches can be used to obtain hierarchical classifications of high granularity in very large collections of research publications at a low cost. This thesis focuses on algorithmic, network-based, approaches to obtain classifications of research publications. I focus on classifications that aim to delineate the publication output of the science system into research fields, which can briefly be explained as focus areas of research, often characterized by a shared subject orientation and relatively dense communication (see Section 3.2).

A great variety of methods have been used during the 60 years of development of algorithmic classifications of research publications. Nevertheless, the procedures generally follow the same steps:

1. Selection of data to be classified.
2. Calculation of publication-publication similarity. This step includes the choice of similarity measure, for example direct citation, co-citation, bibliographic coupling or textual similarity (the measures are explained in more detail in Section 2.1).
3. Normalization of publication-publication relations. For example, to reduce differences between articles that have larger number of citation relations and those that have lower number of citation relations.
4. Choice of clustering algorithm. This step usually includes a choice of parameter values since almost all clustering algorithms have parameters for which users must choose values. Examples of parameters are the number of clusters to be obtained, the number of iterations, or the resolution parameter.
5. Labeling of the obtained clusters. This step may also include parameter settings, for example to balance term frequency with term specificity at different hierarchical levels.

In the remainder of this section, I outline the historical development of algorithmic, citation network-based, classifications of research publications from the 1960s, as well as the current state of this research topic (in the beginning of the 2020s). The focus is on unsupervised, large-scale classification of research publications.

2.1 1960S TO 1980S – EARLY WORK ON NETWORKS OF RESEARCH PUBLICATIONS

It was early recognized that citation databases can be used to study subject orientation of research and development of research topics. Garfield (1963) proposed that citation relations can be used to draw historical maps of science and in 1964 Garfield et al. explored citation relations to study a set of papers associated with "the key discoveries leading to our present understanding of the mechanisms and role of DNA in protein synthesis" (Garfield et al., 1964). They concluded that citation analyses can be used to "identify key events, their chronology, their interrelationships, and their relative importance." At about the same time Doyle (1962) proposed that maps of keywords can be used for information retrieval purposes. A couple of years later de Solla Price (1965b) suggested that research fronts can be studied by drawing networks between journal articles.

Figure 1 shows a citation network between some of the early work covered in this section. Note the contra-intuitive citation from Narin (1972) to the coming work of Carpenter and

Narin (1973). Also note that Marshakova-Shaikevitch (1973) does not have any citation relations with the other early works.

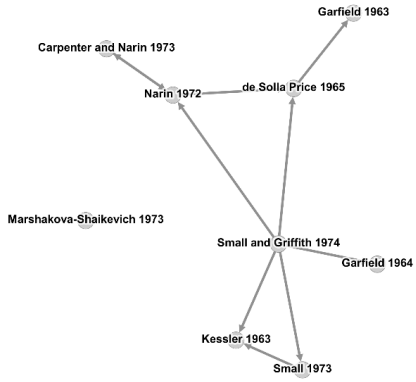

Figure 1: Example of citation network of some of the early work covered in this section.

Different citation-based measures to quantify publication-publication similarity were proposed in the 1950s-70s. In 1954 Fano suggested that the relation between documents can be quantified by "simultaneous references in the literature" (Fano, 1956). Kessler (1963) developed this idea and called it "bibliographic coupling" and two years later he published an experiment of the approach (Kessler, 1965). By bibliographic coupling, the similarity of two publications is quantified by counting the number of references they have in common. Another measure was proposed a decade later independently by Small (1973) and Marshakova-Shaikevich (1973), namely co-citations. The co-citation strength between two publications is the number of times both publications occur in the same reference list in other publications. Direct citations, bibliographic coupling and co-citations are still the most commonly used similarity measures between publications used for unsupervised algorithmic classification.

Figure 2A shows the bibliographic coupling strength between publication A and B. A and B both cite publication C, D and E. Thus, A and B have a coupling strength of three (this strength can be normalized in relation to the total number of references in A and B). Figure 2B shows co-citations between the pairs B-C, B-D and C-D. The co-citations are the result of A citing all of the publications B, C and D.

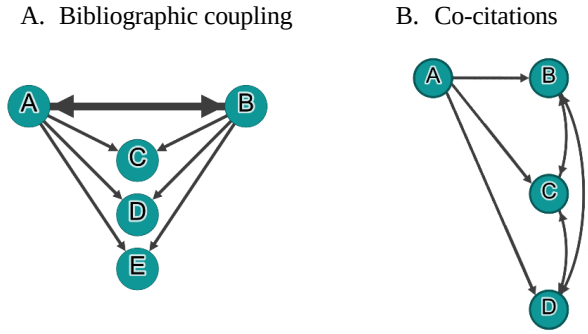

Figure 2: The figure to the left illustrates bibliographic coupling between the publications A and B. Both A and B cite the publications C, D and E. The figure to the right illustrates co-citations between the pairs B-C, C-D and B-D. B, C and D are all cited by the publication A.

After the introduction of bibliographic coupling and co-citations, there has been much discussion and empirical investigation on which similarity measure yields the best groupings of publications into research fields. Later studies have also included other kinds of relations and used other units than publications, such as co-occurrences of words (or terms) in publications (Callon et al., 1991; Callon et al., 1983) and co-authoring between researchers (Persson, 1994; White & Griffith, 1981; White & McCain, 1998).

The first attempts to create network-based classifications of journals started in the 1970s. In 1972 Narin et al. proposed that citations can be used to study the interrelations between journals (Narin et al., 1972) and the year after Carpenter and Narin (1973) partitioned a set of journals in a citation network. They used two different normalization procedures which take into account the different sizes, in terms of number of published articles, of the journals. One of these approaches restricts the network between journals so that for each journal only the top *m* strongest relations are included (setting *m* to 7, 10 and 15). The top *m* approach is of interest because it has also been used in more recent studies at the article level.

Small and Griffith (1974) argued that journals are too broad to be able to capture the structure of science at more finely granular levels. They used co-citations and single-link clustering to create a classification to study the arrangement of publications into research specialties. The single-link clustering method is a type of hierarchical clustering algorithm. In a first step, all relations between publications are calculated. Publications are then grouped step by step. The pair with the lowest distance (i.e., the strongest relation) is joined in each step until all papers have been merged into one cluster. Figure 3 shows an example of a clustering using single-link clustering methodology and direct citation relations between journals in the Web of Science category “Information Science & Library Science”. The single-link clustering methodology has been used in several studies, primarily between 1970 and 2000. A problem with this methodology is that it tends to create chains that merge journals into large groups. This problem was acknowledged by Small and Koenig (1977).

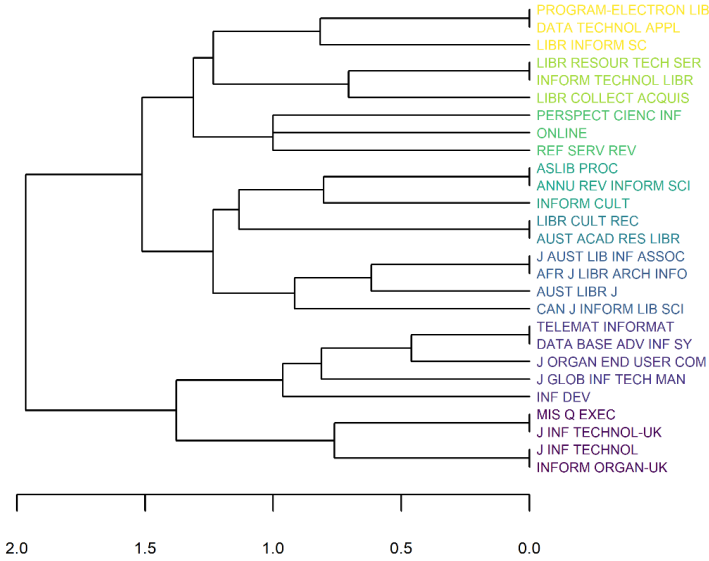

Figure 3: Hierarchical clustering of 25 journals in the Web of Science category "Information Science & Library Science" using direct citation relations and the single-link clustering method. Based on articles from 2015-2019.

Small and colleagues continued to explore clustering using co-citations in the 1980s (Small et al., 1985; Small & Sweeney, 1985). Larger sets of data were clustered using different citation thresholds and different normalization procedures. To prevent the problem of chaining, they proposed a methodology to cut dendrograms at different levels using parameters, including a maximum cluster size.

In 1987 Leydesdorff critically examined the approaches taken to create classifications of science in the previous decades. He pointed out that different clustering methods, different selections of parameters and cut-off points lead to different results and make studies hard to replicate. In particular, he criticized the use of single-link clustering taken by Small et al. (1985). He considered the results to be "an artifact of the applied method". A similar standpoint was taken by Oberski (1988).

In the empirical part of Leydesdorff's paper from 1987, he showed that a clearer clustering structure can be obtained by the use of Ward's clustering method and by changing the similarity coefficient to Pearson's correlation. Ward's clustering method minimizes the within cluster variance. However, it is computationally inefficient because it tries out every possible combination of clusters. Figure 4 shows a clustering using this methodology of the same set of journals as in Figure 3.

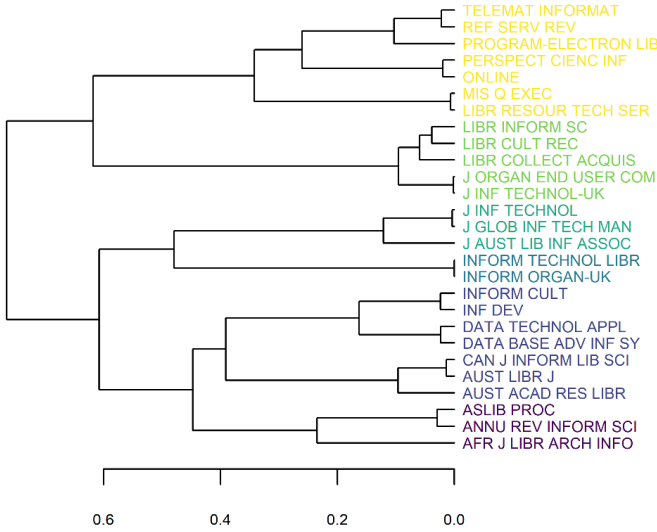

Figure 4: Hierarchical clustering of 25 journals in the Web of Science category "Information Science & Library Science" using direct citation relations and Ward's clustering method. Based on articles from 2015-2019.

Braam et al. (1991) gave support for the results obtained by Small et al. (1985). They compared the word distributions within the clusters and the word distribution external to the clusters. They concluded that clusters do involve coherent research topics and that they "display research specialties, although these may be fragmented into several different clusters."

2.2 1990S AND 2000S – JOURNAL-LEVEL CLASSIFICATION AND NORMALIZATION APPROACHES

Until the 1990s, most clustering attempts had focused on co-citations. Glänzel and Czerwon (1996) pointed out the advantages of using bibliographic coupling over co-citations. They highlighted that "just published papers that are closely related by bibliographic coupling links can provide snapshots of early stages of a specialty's evolution." They made a selection of 4,534 "core documents", publications that are related to other publications with a relatively strong coupling strength. This set was used to study research front topics. An interesting observation is that they found bibliographic coupling to be "sensitive to the length of the chosen publication period." They suggested that bibliographic coupling should be restricted to a relatively short time period and suggested a period as short as 2 years. Such a restriction has, to the best of my knowledge, not been explored in later work.

In the following decade, much effort was put into implementing and improving algorithmic classification at the journal level. The classifications were now scaled to cover all disciplines

in multidisciplinary databases, primarily Web of Science (e.g., see Bassecoulard & Zitt, 1999; Boyack et al., 2005; C.-M. Chen, 2008; Leydesdorff, 2004, 2006).

Leydesdorff (2004) used bi-connected components to cluster a set of 3,991 journals using Pearson's correlation coefficient for normalization of citation relations. This methodology fails to include all connected journals in the final clustering (the initial set consisted of 5,748 journals). The author suggested cosine as a better measure than Pearson's correlation coefficient for normalization because the latter is sensitive to the number of zeros. This property had been criticized in an article by Ahlgren et al. (2003). The cosine normalization approach was taken by Chen (2008), who used affinity propagation for the clustering of 1,578 journals in the Journal Citation Reports.

In 2005 Boyack et al. created maps of 7,121 journals based on direct citation and co-citation relations (Boyack et al., 2005). They used different measures for normalization and restricted the number of relations per journal to the 15 strongest. This restriction keeps the size of the data set down and has also been used and evaluated in several later works at the paper level (Boyack et al., 2011; Boyack & Klavans, 2010; Waltman et al., 2020). A force-directed graph layout algorithm (VxOrd, later called OpenOrd) was used to visualize the networks. K-means clustering (Hartigan & Wong, 1979) was performed to obtain a classification. This clustering method requires the number of clusters to be set beforehand, which is clearly a disadvantage in the context of classification of research publications. Boyack et al. (2005) used Mutual Information to evaluate different normalization procedures, using the Web of Science journal classification as a reference. The authors also highlighted the scalability problem of using Pearson's correlation coefficient for normalization. They preferred the Jaccard index for normalization because of its scalability, its empirical results and qualitative assessment of the visualized networks. A normalization measure that they call K50 performed best when co-citations was used as the journal-journal similarity measure.

Figure 5 shows a network of about 13,000 journals from 2017-2019 using direct citation relations, Jaccard normalization and the OpenOrd layout.

Figure 5: Network of about 13,000 journals from 2017-2019 using direct citation relations, Jaccard normalization and the OpenOrd layout.

In 2008 Rosvall and Bergstrom (2008) proposed a new clustering methodology based on random walks and optimization of the description length of the walker, called the "map equation". They implemented the map equation on a citation network of 6,128 journals, taking the direction of the citations into account. This methodology has gained much attention in other fields, however it has not (yet) been much adapted in scientometric studies (some exceptions are Šubelj et al., 2016; Velden, Yan, et al., 2017; Zeng et al., 2019).

Studies on publication-level classification were of an exploratory nature during this time period, mostly covering small or medium size sets of publications. Different publication-publication similarity approaches were evaluated, both textual approaches and citation-based approaches, as well as the combination of the two (Ahlgren & Colliander, 2009a, 2009b; Ahlgren & Jarneving, 2008; Janssens et al., 2008, 2009; Jarneving, 2007). Clustering methods were also evaluated (Ahlgren & Colliander, 2009a) and different evaluation frameworks were applied, for example by comparing classifications obtained by the use of different methods with classifications created by subject experts (Ahlgren & Colliander, 2009b; Ahlgren & Jarneving, 2008) or the Web of Science journal classification (Klavans & Boyack, 2006). Studies clustering small sets of publications based on full text were also

conducted (Glenisson et al., 2005; Janssens et al., 2006) and hybrid approaches, using both textual similarity and citation-based similarity for clustering, were highlighted in several of the works during the 2000s (Cao & Gao, 2005; Glenisson et al., 2005; Janssens et al., 2009).

2.3 2010S - MODULARITY BASED APPROACHES AND PUBLICATION-LEVEL CLASSIFICATION

The first modularity function was introduced by Newman and Girvan in 2004 and based on that function Newman proposed an algorithm for community detection (Newman, 2004b; Newman & Girvan, 2004). Modularity based approaches for partitioning of nodes in networks into groups have been widely used since. In 2009 and 2010, modularity-based approaches were adapted in the scientometric context and used to cluster publications, journals, authors, and the Web of Science journal classification (Chen & Redner, 2010; Lambiotte & Panzarasa, 2009; Schubert & Soós, 2010; Takeda et al., 2009; Takeda & Kajikawa, 2009, 2010; Wallace et al., 2009; Waltman et al., 2010; Yan et al., 2010; Zhang et al., 2010).

Consider an unweighted network, i.e., a network in which all relations are equally strong. Nodes have been grouped into clusters (or modules or communities) in this network. The modularity of the network is the fraction of the edges that fall within the clusters minus the expected fraction. Modularity based approaches include a quality function and an optimization algorithm. The optimization algorithm tries to find a clustering solution that optimizes the value of the quality function. The number of possible solutions is too large for all to be tested. There is a trade-off between finding the optimal solution and doing it within a reasonable time frame. Many different quality functions and optimization algorithms exist. It is out of scope of this review to extensively cover this literature (for an overview to this historical point in time, see Fortunato (2010)). I will therefore focus on some of the quality functions and optimization algorithms relevant to the scientometric research literature.

The Louvain algorithm was introduced in 2008 (Blondel et al., 2008) and adapted in scientometric work in the following year (Lambiotte & Panzarasa, 2009; Wallace et al., 2009). Compared to previous clustering approaches, the Louvain algorithm is much faster and is able to find clusters in networks of larger scale in reasonable time frames. In 2012, Waltman and van Eck created a large-scale publication-level classification of almost 10 million journal articles using a modified Louvain algorithm (Waltman & van Eck, 2012). The next year, the same authors proposed a new optimization algorithm called the Smart Local Moving algorithm (SLM) (Waltman & van Eck, 2013b). These two articles became a starting point for large-scale publication-level classifications. Figure 6 shows a map of research disciplines based on the approach proposed by Waltman and van Eck (2012) using the more recently proposed Leiden algorithm.

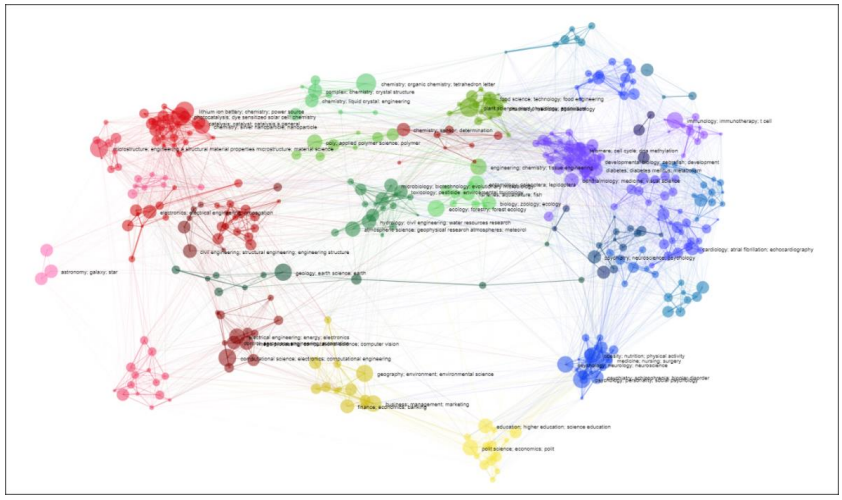

Figure 6: Map of 390 disciplines obtained from about 29 million publications using direct citations and the Leiden clustering algorithm. Biomedicine in blue. Natural sciences in green. Technical sciences in red. Social sciences and humanities in yellow.

Many have argued that there is no one perfect classification and that different similarity measures offer different viewpoints, and the choice of methods should be guided by the purpose of the use of the classification (Glänzel & Schubert, 2003b; Gläser et al., 2017; Klavans & Boyack, 2017a; Mai, 2011; Smiraglia & van den Heuvel, 2013; Velden, Boyack, et al., 2017; Waltman & van Eck, 2012). A special issue in *Scientometrics* explored the use of different methods using one and the same data set ('Same Data—Different Results?', 2017). Gläser et al. (2017) concluded that "we know that several different but equally valid solutions are likely to coexist" and that "[a]lthough we cannot argue by invoking a 'ground truth', we can compare structural properties and contents of clusters."

Much of the literature the last decade has focused on which similarity measure yields clustering solutions that perform best in comparison with different baselines or by the use of different measures (for an overview see Boyack and Klavans (2020)):

1. Calculating the within cluster textual coherence (Boyack et al., 2011; Boyack & Klavans, 2010).
2. Using article-grant groups as a baseline for comparison (Boyack & Klavans, 2010, 2018).
3. Using a set of reference lists of "authoritative papers" as a baseline for comparison (Boyack & Klavans, 2018; Klavans & Boyack, 2017a).
4. Using a text-based measure to evaluate citation-based measures and vice versa (Boyack & Klavans, 2018; Waltman et al., 2020).
5. Comparing clustering solutions with other classifications or controlled vocabulary (Ahlgren et al., 2020; Haunschild et al., 2018).

I agree with others that there is no single answer of which similarity measure performs best because the outcomes of a study are conditional on the choice of baseline. However, the comparisons have contributed to more knowledge about the properties of the similarity measures and the resulting classifications.

It is clear that the co-citation approach excludes all non-cited publications from the classification and thus it cannot be used to cluster recent, non-cited publications. The use of direct citations has a similar problem if small or medium size data sets are used for clustering. If larger sets are used, more publications have at least one citation relation and can thereby be assigned to a cluster in the classification. An extended direct citation approach has been developed to address this problem (Ahlgren et al., 2020; Boyack & Klavans, 2014; Klavans & Boyack, 2017a; Waltman et al., 2020). Let $\boldsymbol{A}$ be a publication set that should be clustered. The extended citation approach considers all publications in $\boldsymbol{A}$ as well as all publications cited by publications in $\boldsymbol{A}$ (but not belonging to $\boldsymbol{A}$). Note that theoretically the direct citation approach equals the extended direct citation approach if the used data source is complete.

The bibliographic coupling approach and the extended direct citation approach have both performed well in several studies (Ahlgren et al., 2020; Klavans & Boyack, 2017a; Waltman et al., 2020). Bibliographic coupling has some properties that are theoretically appealing, which has been acknowledged by others (Glänzel & Czerwon, 1996; Jarneving, 2007). In contrast to co-citations, bibliographic coupling can be used to cluster recent research. Bibliographic coupling also has advantages over direct citations, including extended direct citations. Direct citations indicate a similarity between two publications only in a binary sense, either a citation relation exists, or it does not. The number of bibliographic couplings between two publications indicates if the relation between the publications is strong or weak. A disadvantage with bibliographic coupling is that it is computationally intensive. No study of very large scale has included bibliographic coupling. The largest so far is the study by Ahlgren et al. (2020) including about 3,6 million publications from 2013-2017 from Web of Science. Klavans and Boyack (2020) excluded bibliographic coupling for this reason. Most large-scale classifications have been based on direct citations or extended direct citations. Yun et al. (2020) proposed an approach that splits direct citation networks into citing and cited nodes. This method improves efficiency and results in similar results as co-citations and bibliographic coupling.

Waltman et al. (2020) found that restricting the publication-publication similarity measure to the top m strongest relations increased the accuracy of the clustering rather than lowering it. This is an important observation for efficiency reasons. Ahlgren et al. (2020) did not find that combining citation approaches enhanced performance substantially. This is also an observation that is of relevance for efficiency optimization.

The direct citation approach has not performed as well as bibliographic coupling and extended direct citations in most studies. It should be noted that most studies have restricted data to one or a few research disciplines and a relatively short time period. Boyack and Klavans (2020) got similar results for direct citations and extended direct citations when evaluating different approaches using a data set of about 16 million articles in PubMed from 2000 to 2018 and citations from the NIH Open Citation Collection (NIH OCC). This result indicates that the direct citation approach performs well when classifications are based on large data sets covering a substantial time period and a wide range of disciplines.

Approaches using textual similarity between publications for clustering of publications are more computationally expensive than citation-based approaches. Large-scale implementation may therefore be restricted by the computational capacity. Just a few of the large-scale studies have used textual approaches. Some exceptions are the recent studies by Boyack and Klavans using the PubMed related article measure (RA) (Boyack & Klavans, 2018, 2020; for description of the measure see Lin & Wilbur, 2007). RA is provided by the United States National Library of Medicine. Ahlgren et al. (2020) included the text-based relatedness BM25 measure when evaluating different approaches for clustering of a fairly large publication set of about 3 million publications. The results of Ahlgren et al. (2020) and Boyack and Klavans (2020) give some support for combining textual approaches (RA or BM25) and direct citations. However, Ahlgren et al. (2020) found that the extended direct citation approach outperformed the combined approaches. Extended direct citations also outperformed direct citations combined with RA in the study by Boyack and Klavans (2020) for one of three evaluation approaches (evaluation using references in "authoritative papers").

Evaluation of different clustering algorithms has not been covered to the same extent as evaluation of similarity measures by the scientometric literature. Šubelj et al. (2016) evaluated several clustering algorithms using both statistical properties and expert judgement. Infomap, which is based on the map equation (Rosvall et al., 2009), performed well in this study and was proposed by the authors to be further explored by the scientometric community. The authors point out, however, a disadvantage of Infomap: the obtained classifications have a highly skewed distribution of cluster sizes with a large number of small clusters. This is a potential problem in scientometric applications. The computational time is also higher than for some other approaches, for example the Louvain approach. The authors highlighted that they did not use an optimal resolution parameter value for the modularity-based clustering approach (Louvain). Optimizing the resolution parameter could have made the results different.

There is a lack of other studies evaluating clustering algorithms empirically. However, it should be noted that there are some theoretical arguments which support the use of some methods. Most modularity optimization approaches have a resolution limit (Fortunato & Barthélemy, 2007). Traag et al. (2011) introduced the Constant Potts Model (CPM) to overcome this problem and the same quality function was used by Waltman and van Eck (2012). Since the study by Šubelj et al. (2016), a new modularity-based optimization algorithm has been proposed. This algorithm, called the "Leiden algorithm", is a modified version of the Louvain algorithm and was proposed by Traag et al. (2019). It is shown in the article that the Louvain algorithm sometimes creates badly connected clusters. The Leiden algorithm overcomes this problem and improves connectedness of clusters.

2.4 SUMMARY AND CURRENT STATE

Network approaches for studying research fields and their development started after the emergence of the Science Citation Index (later part of Web of Science) in the 1950s. The most fundamental citation-based measures for publication-publication similarity were

proposed early in this development, i.e., direct citation, bibliographic coupling, and co-citation.

Early work tried to identify specialties by clustering journals. Normalization of the similarity measure was introduced as well as hierarchical clustering methods, such as single-linkage and Ward's clustering. The number of studies grew in the 2000s and new sources emerged. However, the Web of Science was still used in most studies. Several new methods were tested, for example Salton's cosine measure and the Jaccard index were used for normalization, and k-means, OpenOrd and affinity propagation were used for clustering.

Two new clustering methodologies were introduced in the 2000s, modularity-based clustering and map equation. Both have had great influence in network science. Modularity-based clustering was adapted in scientometrics in 2009-2010 and has become the dominant clustering approach.

In the last decade, clustering of research publications has developed towards large-scale implementation at the publication level. More data sources are available and have been used in recent times, for instance PubMed/MEDLINE, Dimensions and the NIH OCC. The extended direct citation approach has been introduced and there has been a focus on evaluating different publication-publication similarity measures.

Table 1 summarizes the historical development of unsupervised, citation-based approaches to classify research publications. The timeline includes seminal work and work of specific relevance in the scientometric context.

Table 1: Timeline over the introduction or scientometric adaptation of different methodological aspects of the first four steps involved in the algorithmic creation of classifications of research publications.

Decade	Step 1: Data source	Step 2: Publication-publication similarity	Step 3: Normalization approaches	Step 4: Clustering and optimization algorithms
1950	ISI Citation Index (later part of Web of Science)			
1960	MEDLARS (origin of PubMed/MEDLINE)	Direct citations (de Solla Price, 1965b) Bibliographic coupling (Kessler, 1963)		
1970		Top *m* approach (Narin et al., 1972) Co-citations (Marshakova-Shaikevich, 1973; Small, 1973) BM25 (Boyack et al., 2011; Robertson & Jones, 1976)	Size of journals (Narin et al., 1972)	Single-linkage (H. Small & Griffith, 1974)
1980			Pearson's correlation coefficient (Leydesdorff, 1987)	Ward's clustering (Leydesdorff, 1987)
1990				
2000	Crossref PubMed Central (PMC) Scopus Microsoft Academic		Salton's cosine (Ahlgren et al., 2003) Jaccard index (Boyack et al., 2005)	Bi-connected components (Leydesdorff, 2004) K-means (Boyack et al., 2005) OpenOrd/VxOrd (Boyack et al., 2005; Davidson et al., 2001) Modularity based approaches (P. Chen & Redner, 2010; Lambiotte & Panzarasa, 2009; Takeda et al., 2009; Takeda & Kajikawa, 2009; Wallace et al., 2009; Waltman et al., 2010) Louvain (Blondel et al., 2008) Complete-linkage (Jarneving, 2007) Affinity propagation (C.-M. Chen, 2008) Map equation (Rosvall & Bergstrom, 2008)
2010	Dimensions OpenCitations NIH Open Citation Collection	PubMed's related article (Boyack et al., 2011; Lin & Wilbur, 2007) Extended direct citations (Boyack & Klavans, 2014; Waltman et al., 2020)	Fractionalization (Waltman & van Eck, 2012)	Modularity Optimizer (Waltman & van Eck, 2012) Leiden algorithm (Traag et al., 2019)

3 CONCEPTUAL FRAMEWORK

In this section, the conceptual framework of the thesis is outlined. In Section 3.1, I discuss the nature of the relation between two publications expressed by a citation. In the following section (3.2), I discuss my view on research fields, including "topics" (3.2.1), "specialties" (3.2.2) and "disciplines" (3.2.3). In section 3.3, I discuss research concepts, focusing on the distinction I make between concepts and topics.

3.1 CITATION THEORIES

Citations play an important role in this thesis. They constitute relations between research publications and create the networks used to cluster publications, argued to represent fields of science at various granularity levels. Hence, it is important to reflect upon the role of citations in communication within research communities and what kind of relation between two publications a citation represents.

A citation is "an expression of a relationship between two documents, the citing and the cited." (Cronin, 1984) This relation is in its basic form binary, it either exists or does not exist. However, the motivations to cite are many and the underlying meaning of citations may differ (Held & Velden, 2022). The binary nature of citations offers a simplified view on the complex ways in which publications are related and on the underlying meaning of the act of citing.

It was acknowledged early on that the motivations to cite a publication varies (Garfield, 1964), including for example "Giving credit for related work", "Identifying methodology, equipment etc.", "Providing background reading", "Criticizing previous work", "Substantiate claims" or "Identifying original publications in which an idea or concept was discussed". Since then, several citation theories have been formulated and the meaning of the relation between a citing and a cited publication has been discussed for a long time period (Aksnes et al., 2019; Amsterdamska & Leydesdorff, 1989; Cronin, 1984; Held & Velden, 2022; Luukkonen, 1997; MacRoberts & MacRoberts, 1989; Robert K. Merton, 1973; van Raan, 1998; Wouters, 1999). This discussion has contributed with different perspectives on the role of citations and the motives to cite (Tahamtan & Bornmann, 2022; Velden, Boyack, et al., 2017; Wouters, 1999).

Two citation theories are dominant in the scientometric literature, the normative theory and the social constructivist theory. The normative theory relates to the norms of science developed by Merton (1973). This theory focuses on citations as acknowledgement of previous work and supports the use of citation indicators to quantify the importance or impact of publications. The social constructivist theory on the other hand focuses on social aspects of citations and emphasizes that citations may be used by researchers for strategic and rhetorical reasons (Aksnes et al., 2019; Tahamtan & Bornmann, 2022).

Tahamtan and Bornmann (2022) have recently proposed a citation theory focusing on citations as communication, rather than motives to cite and their relation to citations as a

measure of impact. The theory is called the social systems citation theory (SSCT) and is based on the social systems theory of Niklas Luhmann. The theory situates citations within the social science system in which communication is the basic constituent element. The SSCT regards research publications to be essential for the communication within the science system and citations as means to separate "own discoveries from knowledge generated by other researchers" (Tahamtan & Bornmann, 2022). Since all relevant publications can rarely be cited, references are typically chosen selectively, and they contextualize and position publications in the science system. The perspective on the role of citations taken in this thesis fits the social systems citation theory in that communication is in focus. The use of citation as a basis to cluster research publications into fields is supported by the selective nature of citations and their function to contextualize research.

3.2 RESEARCH FIELDS

I define a research field as a focus area of research at any granularity level. In this section I primarily focus on research fields at two granularity levels, research topics and research specialties. In addition, I briefly discuss the notion of research disciplines.

Different perspectives on the notions of topics, specialties and disciplines exist (Sugimoto & Weingart, 2015), as well as different perspectives on how to classify science into such units (Hjørland, 1992). In this section, I explain the perspective I take in this thesis, focusing on the formal communication taking place in research publications by referencing, and how this perspective relates to the delineation of science into fields. Other perspectives may capture other aspects of research fields, for example by emphasizing the organization of education, socialization and social networking, language and discourse, or on the use of research literature (Hammarfelt, 2019; Hjørland, 2002; Sugimoto & Weingart, 2015). I would like to emphasize that I do not consider one perspective to be superior. In my view, different perspectives offer insights into different aspects of the properties of the complex structure of fields in the science system, such as how fields emerge, take form, and develop. Using different perspectives provides a more comprehensive understanding of topics, specialties, and disciplines than one would get from taking one perspective only. The perspective taken when classifying research publications into fields must be guided by the intended use of the obtained classifications. Inversely, the choice of classification or other categorization of research publications must be guided by the nature of the problem at hand.

3.2.1 Research topics

I define a research topic (from now on "topic") as a focus area of research that constitutes a thematic context to which researchers relate research questions in research studies. The theme of a topic may be focused to theoretical, methodological, or empirical knowledge. This definition is similar to Havemann et al. (2017), however it does not require topics to be "shared by a number of researchers".

Topics are formed by formal and informal communication between researchers and are constantly created, formed, and recreated within a community or in the overlap of two or

more communities in the science system. Science is an autopoetic (self-reproductive) system (Tahamtan & Bornmann, 2022) and topics are important components of the reproduction of science. Scientific breakthroughs spur the emergence of new topics and the prevailing paradigm in science is sometimes disrupted by revolutionary shifts of focus (Kuhn, 1996). Topics have a relatively homogenous and limited scope but may shift in focus and vocabulary over time. Topics overlap and the same topic can be addressed within different research communities (Yan, Ding, & Jacob, 2012; Yan, Ding, Milojević, et al., 2012).

In this thesis, topics are studied through their presence in research publications. The relation between topics and research publications is important in this context. Original research articles generally address one or a few research questions and contribute to the development of one or a few topics. Review articles often focus on a topic and typically give historical background and current state of the topic (Klavans & Boyack, 2017a). Likewise, PhD theses generally address a topic and contribute with a few studies answering different research questions related to the topic. The topic of a research publication is often mentioned in the background section where related, previous publications are referenced. The findings of a study are usually related to other studies within the same topic in the discussion section. Thereby, researchers relate their publications to other publications within the same topic, which contextualizes their research in the science system.

The scope of a topic is generally not larger than what can be addressed and grasped by a researcher in a literature review. I distinguish between *size* and *scope* of topics. A topic is limited in scope but large in size if containing a high number of publications concentrated to a narrow focus area. Overlap of topics is manifested in publications when two or more topics are addressed in the same publication, for example, when community-detection and science mapping are addressed in scientometric publications.

3.2.2 Research specialties

I define a research specialty (from now on "specialty") as a focus area of research addressed by "*a self-organized network of researchers who tend to study the same research topics, attend the same conferences, read and cite each other's research papers and publish in the same journals.*" This definition is based on the definition by Morris & van der Veer Martens (2008) but differs in that I define a specialty as the focus area addressed by a network of researchers, rather than the network of researchers itself. This makes the definition consistent with being a research field defined as a focus area of research.

In the context of this thesis, it is worth noting the important role of communication within specialties. Researchers within a specialty frequently communicate with each other. The communication of research specialties takes place in shared channels (e.g., journals and conferences) and through referencing. The communication is characterized by shared terminology, knowledge, competencies, and problem areas (Chubin, 1976; Hagstrom, 1970; Scharnhorst et al., 2012). Note that researchers may be part of several specialties and their publications may address topics in different specialties. It is only the events that take place

within the context of the specialty that is part of the specialty. In similarity with topics, specialties are focus areas of research. However, specialties are broader in scope and range over a set of topics related to the specialty. Specialties are the largest somewhat homogenous communities in the science system (Colliander, 2014).

3.2.3 Research disciplines

Research disciplines (from now on "disciplines") are not easily defined and identifying the publications related to a discipline is difficult. Characteristics involved in the identification of a discipline may include: (a) the organizational structure of research bodies (in particular universities), (b) a set of shared methods, theories, and concepts, (c) shared epistemology, (d) researchers with shared educational background and specialization, and (e) a shared problem area (Hammarfelt, 2019). Different disciplines can be defined by different types of characteristics. For example, "Pediatrics" is defined by the groups of people in focus for medical care (infants, children, and adolescence), "Urology" by the part of the human body in focus (the urinary system) and "Obstetrics" by a process (pregnancy and childbirth). Disciplines are not in focus for this thesis. I have chosen to use disciplines to denote the clusters resulting from clustering specialties. This may not coincide with definitions used by others (for an overview, see Sugimoto & Weingart, 2015). The choice has simply been guided by the fact that the labeling procedure at this level has resulted in labels frequently including terms that are commonly understood as disciplines, for instance "chemistry", "biology", "immunology", "rheumatology" or "anesthesiology". I recognize that the denotation of the clusters at this level has not been thoroughly investigated and that other definitions of disciplines would lead to other operationalizations.

3.3 RESEARCH CONCEPTS

In this section I discuss the distinction I make between research topics and research concepts (from now on "concepts"). A wider discussion about concepts is out of scope of this thesis and I restrict this definition to concepts as manifested in research publications.

In analog with the MeSH system, I consider a concept as a bearer of linguistic meaning related to an aspect of a research publication.[9] Each MeSH category corresponds to exactly one concept and synonyms refer to the same concept.[10] MeSH categories are good examples of concepts; however, MeSH is not complete and other structures of concepts are possible. Concepts are generally used to be able to search or delimit search results in an information retrieval system.

The scope of a concept varies, some are broad, and some are narrow. A concept of narrow scope describes an aspect of a research publication in high detail (e.g., "smooth muscle myocytes"). Such a concept is situated on a low hierarchical level, having none or few

[9] "Concept Structure in MeSH". https://www.nlm.nih.gov/mesh/xml_data_elements.html#Concept [2021-11-26]
[10] "MeSH XML Data Elements". https://www.nlm.nih.gov/mesh/concept_structure.html [2021-11-26]

underlying concepts. A concept of broad scope describes a more general aspect of a research publication without much detail (e.g., cells). A concept of narrow scope is large in size when a lot of research activity is related to the concept.

Concepts can be of different *type*. This is well exemplified by the structure of the MeSH tree, where the major categories contain concepts ordered into physical entities (e.g., “Anatomy”, “Organisms” and “Chemical and Drugs”) and others into different processes or activities (“Phenomena and Processes” and “Health Care”). Other examples are “Geographicals”, “Disciplines and Occupations”, “Named Groups” and “Publication Characteristics”.

In contrast to topics, concepts do not necessarily correspond to focus areas addressed by researchers. They may, for example, be broader in scope. To illustrate this, we may take the concept “diet”. “Diet” is related to different disciplines, such as biochemistry, dentistry, psychiatry, oncology and public health. A search in PubMed on the MeSH term retrieves about 300 thousand publications. The concept is far too broad to correspond to a topic, as defined in this thesis. However, if combined with other concepts such as “vegan” and “dental health” we may identify a focus area that could be considered a topic, which we may entitle as “the association between vegan diet and dental health”. This also illustrates the suitability of concepts to be used as labels of research topics.

4 DATA AND METHODS

In this section, I describe the bibliographic data sources used in the thesis and the general properties of citation networks. I then introduce the clustering methodology used throughout the thesis. Data and methods that are specific to the individual studies are presented in the results section in the summary of the article of relevance.

4.1 BIBLIOGRAPHIC DATA SOURCES

Two primary data sources have been used for the studies in this thesis: the Web of Science and PubMed/MEDLINE in conjunction with the NIH OCC.[11] Web of Science is a multidisciplinary bibliographic database owned by Clarivate Analytics.[12] It has a higher coverage in medicine, natural sciences and engineering and a lower coverage in the social sciences and humanities (Martín-Martín et al., 2021). PubMed/MEDLINE is maintained by the National Library of Medicine (NLM) in the United States. It has a broad coverage of biomedicine but does not include references, nor citation relations, in the bibliographic records. I have used the NIH OCC to complement the PubMed/MEDLINE records with citation relations. The NIH OCC is restricted to citation relations within PubMed/MEDLINE. Both PubMed/MEDLINE and NIH OCC are openly available. The coverage of citation relations in NIH OCC has been growing immensely during the work with this thesis, as the result of more references becoming openly available in Crossref during the time period (Hutchins, 2021; Martín-Martín et al., 2021; Visser et al., 2021).[13] Presently (December 2022), more than 90% of the articles from the latest 10-years period in PubMed/MEDLINE have references in NIH OCC.

4.2 CHARACTERISTICS OF CITATION NETWORKS

The citation relations within the publications covered in the two data sources have been used to create giant networks of publications – the publications being the nodes of the network and the citations being the edges. The networks used for clustering in Articles I-IV include 17-31 million publications, having 400 million to 1 billion citation relations. In this section, I outline the main characteristics of these networks. The main characteristics described in this section are general to citation networks and are not likely to be dependent on data source.

Citations are by their nature *directed* relations between publications, i.e., there is a source from which the citation "emanates" and a target where it is "received". Citations emanate from newer publications to older publications and, thus, have a directionality backwards in time (a few contra-intuitive exceptions can be found as the citation in Figure 1 from Narin (1972) to Carpenter and Narin (1973)). There is a tendency to refer to recent literature more

[11] See individual papers for specifications on included indexes and other restrictions.
[12] Web of Science data included in this thesis are derived from the © Web of Science 2023 of Clarivate Analytics (UK) Ltd. All rights reserved.
[13] Crossref is a not-for-profit membership organization maintaining a bibliographic data source that interlinks various data related to publication meta data.

frequently than to older literature. The bars in Figure 7 show how references from the publication year 2020 are distributed over the time period of 1995-2020. The blue line indicates the expected number of citations to each year.[14] The figure shows that there is an excessive number of citations from 2020 to the previous 5-10 years with a peak about 2-3 years before the citing publication date.

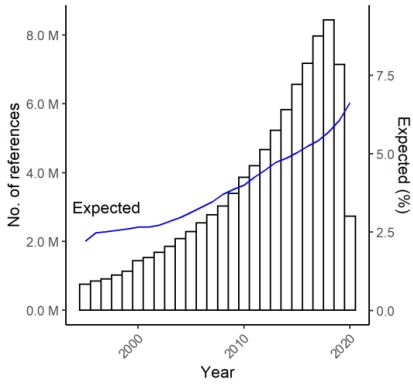

Figure 7: Distribution of references from 2020 over publications published 1995-2020 (Web of Science).

Direct citation networks of research publications are extremely sparse, i.e., only a very small proportion of the possible edges exists in the networks. To indicate the sparseness, we can use the number of publications in Google Scholar as a proxy for the maximum possible number of references in an article. Gusenbauer (2019) estimated the number of publications in Google Scholar to be around 400 million in 2019. Thus, an article written in 2020 could theoretically include roughly 400 million references. However, articles published in 2020 only included about 49 references on average, a fraction of about 1.2×10^{-7}. The sparseness comes as no surprise; it would neither be realistic, nor is it allowed by journals, to refer to a large proportion of all existing research literature in a single publication.

Table 2 shows summary statistics of the number of references for articles published from 2016 to 2020 and registered in Web of Science. The number of references increases over time. In 2020 most articles (10^{th} to 90^{th} percentiles) include about 20-80 references. As shown by Figure 8, the distribution of the number of references in publications is approximately normal, but it has a small excess of zeroes and a small tail to the right. Thus, a small share of publications has more references than one would expect from a normal distribution (up to thousands in some cases).

[14] The expected proportion of citations to year *y* from 2020 is defined as:

$$E_y = \frac{P_y}{P}$$

where P_y is the number of publications in year *y* and *P* the total number of publications published in 1995-2020.

Table 2: Summary statistics of the number of references in articles published in the years 2016 to 2020 (Web of Science).

Year	N	Sum	Mean	Min	P10	Q1	Median	Q3	P90	Max
2020	2,149,547	105,357,700	49	0	19	29	42	59	83	5,718
2019	1,899,045	88,920,279	46.8	0	18	28	40	56	79	2,945
2018	1,783,757	81,171,860	45.5	0	18	27	39	55	77	8,674
2017	1,695,914	74,741,060	44.1	0	17	26	37	53	75	4,864
2016	1,636,291	70,045,038	42.8	0	16	25	36	52	73	8,371

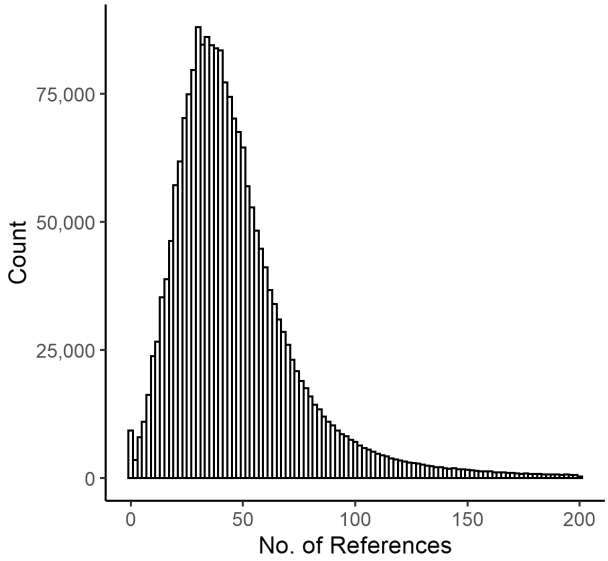

Figure 8: Distribution of number of references per publication in the year 2020. Delimited to publications with 200 references or less for readability of the graph.

If we turn to the receiving side of the citation relations, a very different distribution emerges compared to the distribution of references. The distribution of citation counts over publications is highly skewed, with an excessive number of publications receiving no or just a few citations and a low proportion receiving a moderate to high number of citations (Figure 9). Most publications (95%) do not receive more than 50 citations in a five-year period (publications from 2015, counting citations until April 2021 in Web of Science). However, some publications receive thousands of citations.

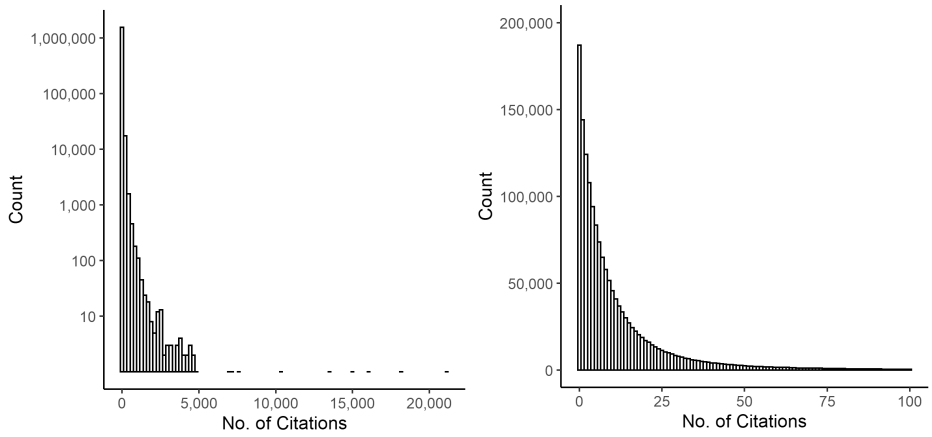

Figure 9: Distribution of number of citations per publication in the year 2015. Citations counted until April 2021. Left histogram with log10-scale on the y-axis. Right histogram delimited to publications with a maximum of 100 citations.

From these data we get a picture of a typical publication in the citation network. It includes around 30-50 references. A few of these references are expected to point to highly cited publications, but most point to publications with one or a few citations. The publication itself is not likely to have more than a few citations.

Another important property of the citation network is the varying density of citation relations within different disciplines. Table 3 shows summary statistics of citation counts for publications published in 2015 for some journal categories in Web of Science. The average number of citations varies from about 0.2 in "Literary Reviews" to 28 in "Nanoscience & Nanotechnology". This variation is due to several factors. First of all, the average number of references varies between fields. Secondly, the coverage of publications varies between fields and data sources. For example, Web of Science has a low coverage of publications in social sciences and humanities and disciplines within these fields have lower citation counts for this reason. Yet another explanation of the varying citation counts between fields is the age of the referenced literature. Citations are always counted within a restricted period of time. For example, if we count citations for publications published in 2015 until the point in time when these data were retrieved (April 2021) we cannot count citations for more than approximately 5 years. Fields that generally refer to new literature, not older than 5 years, will have higher citation counts after a 5-year period in comparison with publications in fields referencing to older literature.

Table 3: Summary statistics of the number of citations in articles published in 2015 by Web of Science journal category. Restricted to some journal categories for exemplification.

Journal category (Web of Science)	N	Mean	Min	P10	Q1	Median	Q3	P90	Max
Nanoscience & Nanotechnology	34,487	28.2	0	1	4	12	30	66	3,184
Engineering, Environmental	13,165	25.0	0	2	6	14	30	56	1,266
Oncology	43,596	20.8	0	2	5	11	22	42	16,122
Materials Science, Multidisciplinary	92,264	20.4	0	1	3	8	21	46	3,184
Biochemistry & Molecular Biology	55,111	19.5	0	1	4	9	20	40	7,648
Engineering, Chemical	31,930	17.1	0	1	3	8	19	38	1,354
Business	7,017	15.7	0	1	3	8	19	38	2,320
Computer Science, Artificial Intelligence	12,885	14.3	0	0	2	6	15	32	4,539
Psychology	7,290	13.7	0	1	3	7	16	29	739
Urban Studies	2,275	12.4	0	1	2	6	14	29	401
Development Studies	2,170	10.2	0	0	2	5	12	24	243
Information Science & Library Science	4,054	10.0	0	0	1	4	11	23	670
Sociology	5,750	8.3	0	0	1	4	10	19	541
History	7,257	1.1	0	0	0	0	1	3	59
Literary Reviews	2,056	0.2	0	0	0	0	0	0	64

4.3 CHARACTERISTICS OF NETWORKS BASED ON BIBLIOGRAPHIC COUPLINGS AND CO-CITATIONS

In contrast to direct citations, bibliographic couplings are undirected. Because of the tendency to refer to recent publications (as shown in Figure 7), bibliographic couplings are more likely to occur between publications published close in time. Note that a publication must be cited at least twice to generate a bibliographic coupling. Hence, all references to publications receiving only one citation are disregarded. Consequently, a publication must have references to publications with at least two citations for it to be bibliographically coupled with other publications.

The number of couplings generated by a cited publication increases quadratically by the number of citations to the publication. To exemplify, a publication with 10 citations generates 45 couplings (10*(10-1)/2), a publication with 100 citations generates 4,950 couplings and one with 1,000 citations generates 499,500 couplings. Because of this property, highly cited publications generate a large proportion of the bibliographic couplings in citation databases.

Some of the characteristics that apply to bibliographic couplings also apply to co-citations. Co-citations are undirected and more likely to occur between publications published close in time. There must be at least two references in a publication for a co-citation to occur. Highly cited publications generate many relations between themselves and other publications. This contrasts with bibliographic coupling for which highly cited publications generate relations between the citing publications. Only publications that have been cited have co-citations. Thus, new publications do not have co-citations before they have been cited. Therefore, it is not possible to cluster new publications by the use of co-citations before they have been cited.

There are usually fewer co-citations than bibliographic couplings in citation networks. This is because of the skewed distribution of citation counts and the more normal distribution of reference counts. To exemplify, for a set of about 36 million publications in Web of Science of the publication types “article” and “review” and the time period 1995-2022, there is about 162 billion bibliographic couplings, 22 billion co-citations and 0.85 billion direct citations. 1.2 million of the publications have no direct citation relations, 3.2 million have no bibliographic couplings and 5 million have no co-citations. This illustrates that even though the direct citation approach results in fewer edges, a higher proportion of the publications in the data source can be included in a direct citation network than networks based on bibliographic coupling or co-citations. It should be noted that direct citations, bibliographic coupling, and co-citation are all based on the same data.

4.4 NORMALIZATION OF PUBLICATION-PUBLICATION RELATIONS

There are several reasons to use normalization of edge weights in citation networks. The first is the differences in edge density in different research fields, caused by the variation in the number of references, database coverage and the age of the referenced work, as discussed above. Another reason for normalization is the highly skewed distribution of citations over publications. If no normalization is performed, highly cited publications influence the results of the clustering heavily. This may lead to unwanted properties of the clustering solution, such as a highly skewed distribution of cluster sizes. Yet another reason for normalization is the differences in age of the publications. Older publications have had more time to receive citations than newer publications. Thus, they will generally have more citation relations. This property will cause a bias in the citation network, leading to a higher influence of older publications on the clustering results. Depending on the application, such bias may be unwanted.

I have used the normalization approach proposed by Waltman and van Eck (2012) in the studies of the thesis. I have chosen to denote this approach by the term “fractional

normalization approach". In this approach, the normalized edge weight of publication *i* with publication *j* is defined as:

$$a_{ij} = \frac{c_{ij}}{\sum_k c_{ik}} \quad (1)$$

where c_{ij} is 1 if *i* cites *j*, 0 otherwise and $\sum_k c_{ik}$ the total number of relations of *i*. I have treated the citation networks as undirected when clustering. For a relation between *i* and *j*, where *i* cites *j*, a_{ji} has been calculated equivalently as a_{ij}. The average of a_{ij} and a_{ji} has then been used as the weight of the edge.

4.5 MODULARITY OPTIMIZATION

The modularity of a cluster is given by comparing the observed number of edges within the cluster with an expected value based on randomly distributed edges (Newman & Girvan, 2004). For weighted graphs, the sum of edge weights is used in the calculation. Modularity-based approaches maximize a quality function. Different quality functions and optimization algorithms exist (Newman & Girvan, 2004; Reichardt & Bornholdt, 2006; Traag et al., 2011; Waltman & van Eck, 2012). I have used the quality function given in Waltman and van Eck (2012) in the first two articles of the thesis and the Constant Potts Model (CPM) (Traag et al., 2011) in the two last articles. The two quality functions are equivalent and therefore I only present the CPM. CPM is defined as:

$$\mathcal{H} = -\sum_{ij}(A_{ij}w_{ij} - \gamma)\delta(\sigma_i\sigma_j) \quad (2)$$

where $A_{ij} = 1$ if there is a citation relation between *i* and *j*, w_{ij} the weight of the citation relation and γ a resolution parameter. σ_i is the cluster of *i* and σ_j the cluster of *j*. If $\sigma_i = \sigma_j$ then $\delta(\sigma_i\sigma_j) = 1$, otherwise zero. Desirable clustering solutions minimize the value of $\mathcal{H}$.

Some previous quality functions were not able to detect small communities. This problem is known as the "resolution limit problem" (Fortunato & Barthélemy, 2007). CPM is resolution-limit-free (Traag et al., 2011). The γ-parameter is used to adjust the granularity of the cluster solution. Thereby, CPM can be used to obtain clusters at different granularity levels.

Until a couple of years ago, the most popular modularity optimization algorithm was probably the Louvain algorithm (Blondel et al., 2008). In short, the Louvain algorithm starts by assigning each node to a cluster. It then iterates through all nodes. For each node, the algorithm calculates the change of the quality function if moving the node to the cluster of its neighbor nodes. The node is assigned to the cluster resulting in the highest positive increase of the quality function. The network is then aggregated so that each partition is considered a node in the network. Both phases are repeated until the value of the quality function cannot be further optimized.

Traag et al. (2019) showed that the Louvain algorithm is associated with some problems. In particular the Louvain algorithm creates clusters that are badly connected or even internally disconnected (i.e., clusters with at least two nodes not connected by a path). The Leiden

algorithm introduces a refinement phase to overcome this problem and guarantees that each cluster is internally connected. The clusters obtained by the Leiden algorithm are internally better connected than the clusters obtained by the Louvain algorithm, which makes the Leiden algorithm preferred.

Modularity optimization approaches for clustering have been used throughout the thesis. Other clustering approaches exist, such as hierarchical clustering approaches, flow-based approaches and k-clustering (Fortunato, 2010). In scientometrics, modularity-based approaches have in recent years been the most commonly used approach for clustering. This is partly for efficiency reasons. However, it should be recognized that there is, to the best of my knowledge, a lack of studies evaluating clustering approaches. Evaluation of different approaches is out of scope for this thesis. However, the methods proposed in Articles I-IV are applicable to other clustering approaches.

5 RESULTS – SUMMARY OF ARTICLES I-IV

In this section, I summarize the articles of the thesis. Articles I and II address the granularity of classifications in relation to topics and specialties. Article III addresses labeling of clusters of different granularity and Article IV addresses visualization of the obtained classifications, incorporating coarse clusters for overview and small clusters for detail.

5.1 ARTICLE I – "GRANULARITY OF ALGORITHMICALLY CONSTRUCTED PUBLICATION-LEVEL CLASSIFICATIONS OF RESEARCH PUBLICATIONS: IDENTIFICATION OF TOPICS"

In Article I, we investigated how clusters corresponding to topics can be obtained. We constructed a set of baseline classes based on a set of publications and their references. We then obtained clustering solutions with different granularity. Each clustering solution was compared to the baseline classification and the clustering solution showing the highest similarity with the baseline classification were used to study topics in two case studies.

5.1.1 Data and methods

To create a baseline, we started by retrieving publications with more than 100 references. Publications with more than 100 references are likely to be authored by researchers with great expertise, be review papers, have high impact and summarize a topic (Klavans & Boyack, 2017a). Therefore, we argue that each such baseline publication can be used as a proxy for a topic. We considered each baseline publication as a class and its references as the members of the class.

We used about 31 million Web of Science records from the in-house bibliometric system at KTH Royal Institute of Technology for the analysis. We restricted the set of baseline publications to those having at least 80% active references, i.e., references pointing to other publications in the data set. To avoid having more than one class within a topic we created groups of publications having a citation overlap of at least 30% of their references. From each such group of connected components we selected one publication randomly. After this procedure we made sure that each referenced publication was included in exactly one baseline class. We assigned publications belonging to more than one baseline class to the class to which it had the highest bibliographic coupling strength.

When the baseline classes had been created, we obtained several cluster solutions of various granularity. Each such cluster solution was compared to the baseline classes using the Adjusted Rand Index (ARI). We propose that the clusters at the granularity level with the highest ARI value can be used to approximate the scope of topics. We denote the best performing cluster solution as $ACPLC_t$ and argue that each cluster within this solution roughly corresponds to a topic.

5.1.2 Results

Most publications in $ACPLC_t$ belong to clusters having a size of 70-700 publications (10^{th} to 90^{th} percentile in the weighted cluster size distribution), with a yearly output of about 5-80 publications. The yearly output has likely increased since the time of analysis and we can expect the current output to be around 10-100 for corresponding clusters. However, it should be noted that both coarser and more granular cluster solutions had similar ARI values, so the size of topics should be considered as approximations.

To evaluate if $ACPLC_t$ can be used to study topics, we provide a couple of case studies. The first being topics in Journal of Informetrics. We analyzed the distribution of 632 articles into the clusters of $ACPLC_t$. The topics identified include the following: (1) researcher-level indicators, (2) normalization of bibliometric indicators, (3) science-mapping, (4) research collaboration and (5) measures of publication-publication similarity. The second case study focused on nano-cellulose materials. We used an extensive review article covering 391 references, of which 227 can be found in our clustering solution (most of the excluded references were to conference papers and patents). The review article delineates the research on nano-cellulose materials into three groups based on the different methodologies used to obtain such materials. We calculated the distribution of references over clusters and found that most (about 75%) were concentrated into three clusters corresponding to the three groups mentioned in the paper. The rest of the references were distributed into many different clusters.

We conclude that the proposed methodology may give guidance of how to set the value of the resolution parameter of the clustering algorithm in order to obtain clusters that can be used to detect and study topics.

5.2 ARTICLE II – "GRANULARITY OF ALGORITHMICALLY CONSTRUCTED PUBLICATION-LEVEL CLASSIFICATIONS OF RESEARCH PUBLICATIONS: IDENTIFICATION OF SPECIALTIES"

Building on the results from the first article, we aggregated the obtained clusters at the topic level into larger clusters in Article II. We used a baseline of a set of journals and their articles to calibrate the resolution parameter of the clustering algorithm. Thereby, we obtained clusters that can be used to detect and study specialties.

5.2.1 Data and methods

We based the analysis on the same set of about 31 million Web of Science publications and the cluster solution obtained in the first study ($ACPLC_t$). We restricted the set of baseline journals with the intention to retrieve journals focused to specialties. As a first step we restricted the time period to 2008-2012, decreasing the risk to include journals shifting in focus over time. Next, we removed journals in the Web of Science category "Multidisciplinary Sciences". We also excluded very small or very large journals by including journals between the 10^{th} and 75^{th} percentiles in the size distribution. As a result, journals publishing 47-478 publications in the time period were included. With the intention

to exclude journals having a focus broader than specialties, we restricted the set to journals with at least 10% journal self-citations. To avoid inclusion of multiple journals within the same specialty, we used bibliographic coupling to create groups of journals with similar subject orientation. A threshold of 8% was used to consider two journals as overlapping and we then selected one random journal from each connected component. The final set of baseline journals consisted of 967 journals. We considered each journal as a class and the publications published in the journal in the given time period as the members of the class.

We used the same procedure as in Article I to identify a cluster solution that best matches the baseline. Several clustering solutions were obtained using different resolution parameter values. Each solution was compared to the baseline using ARI. The solution with the highest ARI value was denoted $ACPLC_s$. We argue that the clusters of $ACPLC_s$ roughly correspond to specialties.

5.2.2 Results

Most publications in $ACPLC_s$ belong to clusters having a size of about 3,000 to 23,000 publications (10^{th} to 90^{th} percentile in the weighted cluster size distribution). Similar to the topics study, other clustering solutions performed almost equally. Hence, the size of specialties are rough estimates. Taking this into account, we can conclude that specialties generally range in size from about 2,000 to 40,000 publications, considering publications from 1980-2017 in Web of Science, with a yearly output around 100 to 2,000 publications.

To evaluate if the results make intuitive sense, we explored two Web of Science journal categories and how publications from 2011–2015 in these categories were distributed over clusters. The first category was “Information Science & Library Science”. The largest cluster in this category was a scientometrics/bibliometrics cluster, followed by a library science cluster, a cluster specific to academic libraries and an information retrieval cluster. The second journal category was “Medical Informatics”. The cluster with most publications in this category addressed medical informatics in clinical settings, including electronic health records. The second cluster addressed online health and the third largest cluster focused on health technology assessment.

We conclude that by calibrating the resolution parameter of the clustering algorithm, clusters are obtained corresponding to the size of specialties. The results of this study indicate that labeling is an even harder problem at this level, than the more granular level of topics.

5.3 ARTICLE III – “ALGORITHMIC LABELING IN HIERARCHICAL CLASSIFICATIONS OF PUBLICATIONS: EVALUATION OF BIBLIOGRAPHIC FIELDS AND TERM WEIGHTING APPROACHES”

In Article III, we address the problem of labeling clusters of publications. This was done by constructing an evaluation framework based on MeSH terms and Science-Metrix Journal Classification (SMJC). We evaluated different bibliographic fields and term weighting

approaches at various granularity levels. Based on this evaluation, we provide three recommendations for labeling clusters at different levels of granularity.

5.3.1 Data and methods

Web of Science data from the bibliometric database at Karolinska institutet was used for the analyses. As in the two previous studies we created baselines in order to compare different labeling approaches. A first baseline was created using MeSH. Each MeSH term was considered a class and all publications assigned the MeSH term as "major descriptor" were considered as the members of the class. We argue that the MeSH term is an accurate label for the corresponding class. Thereby, we obtained a baseline of classes (called MeSH classes) that can be used to evaluate labeling approaches. A second, coarser, baseline was created using the SMJC. Each category in SMJC was considered a class and the publications in the journals of each category as the members of the class. We argue that the category name is a suitable label for each class.

To obtain labeling candidate terms, we extracted noun phrases from different bibliographic fields. We operationalized noun phrases as sequencies of adjectives and nouns ending with a noun. Five bibliographic fields were used: (1) titles, (2) abstracts, (3) author keywords (referred to as "keywords"), (4) journal names and the (5) suborganization field in the author addresses (referred to as "addresses").

Next, we used different approaches to calculate the relevance of the candidate terms to baseline classes. They all have in common that they balance the frequency of terms in classes and the specificity of the terms to each class. However, the approaches balance frequency and specificity in different ways. The following approaches were evaluated: Chi-square, Jensen Shannon Divergence (JSD), term frequency-inverted document frequency (TF-IDF) and an approach used in Waltman and van Eck (2012), which we denoted as WvE. The WvE approach was tested using different values of a parameter (m). Higher values of m give more weight to frequency and lower values more weight to specificity. We also proposed a new approach which we denote as "term frequency to specificity ratio" (TFS). TFS is the weighted geometric mean of term frequency and specificity. The weights of frequency and specificity can be adjusted by a parameter (α). Higher values give more weight to frequency and lower values give more weight to specificity.

We evaluated combinations of bibliographic fields and term weighting approaches using Match@N (Carmel et al., 2009; Mao et al., 2012; Treeratpituk & Callan, 2006). This indicator considers whether any of the top N terms can be found in the baseline. All evaluated combinations of bibliographic fields and term weighting approaches were compared using Match@N using N=3. This was done both using the MeSH baseline and the SMJC baseline.

5.3.2 Results

Using a combination of phrases from titles and keywords resulted in the best match between algorithmically obtained terms and the labels in the MeSH baseline (Figure 10). Chi-square

performed best of the term weighting approaches. However, several other approaches performed almost equally good, in particular TFS (with α=0.33 and α=0.5) and WvE with low values of *m*.

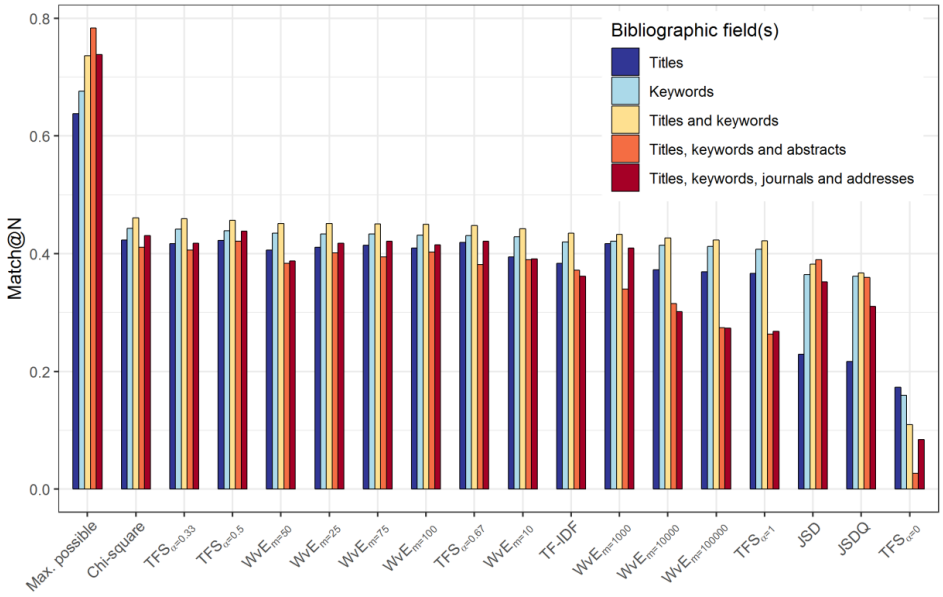

Figure 10: Match@N rates of different combinations of a term weighting approach and one or more bibliographic fields. Match@N rates are based on the MeSH baseline. Approaches are ranked in descending order of their Match@N rate obtained using titles and keywords.

Using the SMJC baseline, the best result was obtained for the combination of journal names, addresses, titles, and keywords (Figure 11). However, the combination of journals and addresses resulted in almost as high Match@N values. JSDQ (a version of JSD), performed best among the term weighting approaches, but was followed by several approaches with similar Match@N levels, namely TFS (with α=0.67 and α=0.5), TF-IDF and WvE with high values of *m*.

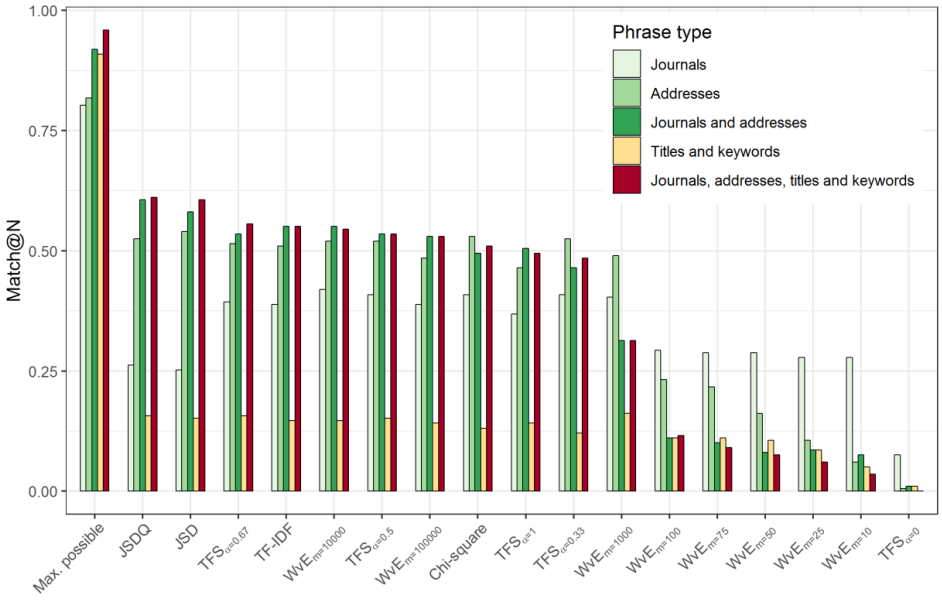

Figure 11: Match@N rates of different combinations of a term weighting approach and one or more bibliographic fields. Match@N rates are based on the SMJC baseline. Approaches are ranked in descending order of their Match@N rate obtained using journals, addresses, titles, and keywords.

We conclude that terms from titles and keywords are suitable to label clusters at the topic level. For most of the evaluated approaches, abstracts were found to decrease the Match@N rates. At higher levels of granularity, we propose terms from journals and addresses to be used for labeling. Furthermore, we suggest TFS to be used to calculate term relevance at various granularity levels. TFS performed reasonably well in both baselines using α=0.5, i.e., giving equal weight to frequency and specificity. The alpha parameter can also be adjusted to lower values at more granular levels and higher values at coarser levels.

5.4 ARTICLE IV – "IMPROVING OVERLAY MAPS OF SCIENCE: COMBINING OVERVIEW AND DETAIL"

In Article IV, I synthesize the previous papers of the thesis by obtaining a hierarchical classification based on the knowledge resulting from these previous studies. In addition, I propose how hierarchical classifications of research publications can be visualized to provide both overview and detail of the biomedical sciences.

5.4.1 Data and methods

A total of about 17 million publications from PubMed and about 380 million citation relations from the NIH OCC were used in this study. The publications were clustered into three levels of aggregation. At the most granular level – topics – the choice of resolution parameter value was guided by the results from Article I. The obtained clusters were

aggregated into larger clusters – specialties – using a resolution parameter value guided by the results of Article II. A third level was obtained by adjusting the resolution parameter in order to obtain approximately the same number of clusters as the number of categories in well-known journal classifications. I denote the clusters at this level as “disciplines”. Labels were created building on the results from Article III. Noun phrases were extracted from titles, journal names, and addresses. Since MeSH terms were available for the publication set, I chose to extract noun phrases from the publications’ MeSH terms, rather than using the author keywords as in the previous study. An illustration of the steps taken to obtain the classification and label the clusters is given in Figure 12, steps 1-5.

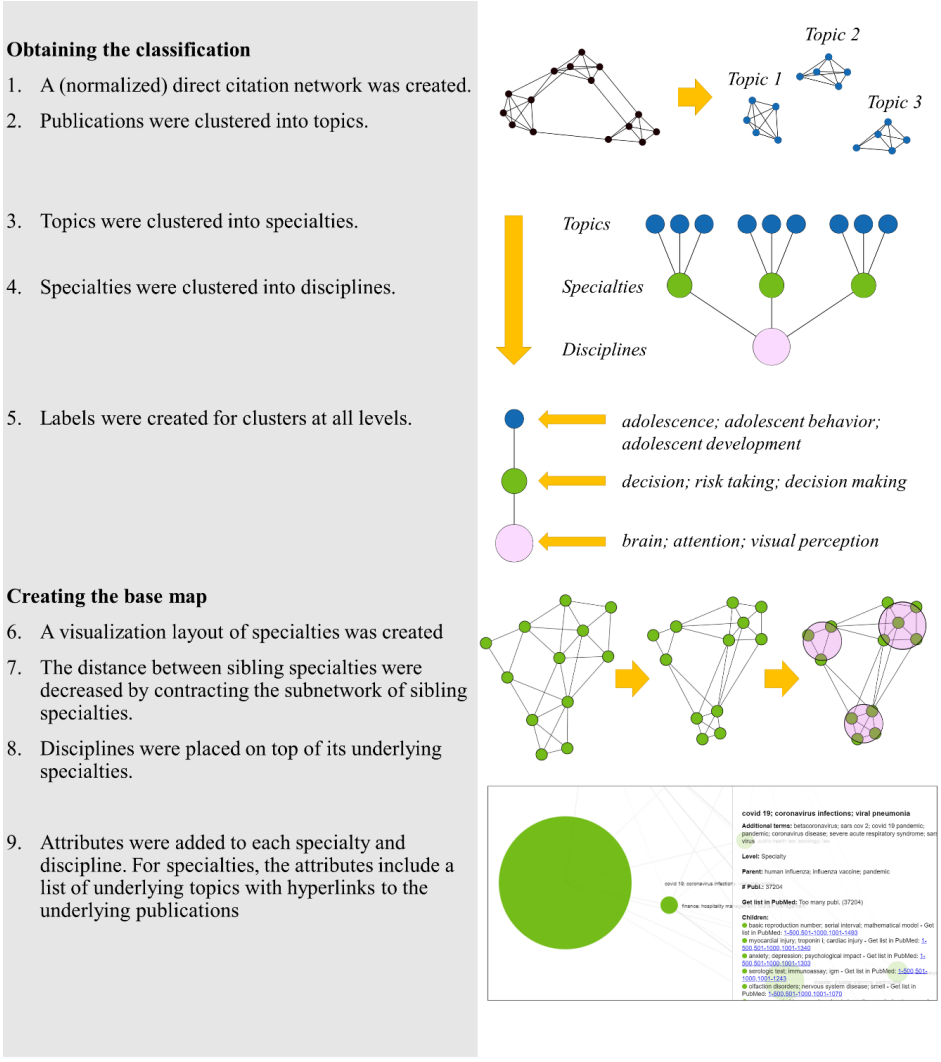

Figure 12: Illustration of the process used to create and visualize the classification.

Steps 6-9 in Figure 12 illustrate the approach used to visualize the obtained classification. A layout was created of specialty-level clusters based on their normalized citation relations (step

6). The ForceAtlas algorithm was used to create the layout (Jacomy et al., 2014), but I acknowledge that other layout algorithms may be used (Fruchterman & Reingold, 1991; Kamada & Kawai, 1989; van Eck et al., 2010). To emphasize the hierarchy of the classification, sibling specialties (specialties clustered into the same discipline) were contracted in order to position them in proximity. A parameter was used for this purpose, making it possible to employ more or less contraction of sibling specialties. When the final layout of specialties had been obtained, I positioned each parent discipline in the center of its set of underlying specialties, given by the average x and y coordinates. Note that the positions of disciplines are not affected by the contraction of specialties.

The methodology was illustrated by two case studies using overlay maps (Rafols et al., 2010), defined as "base maps over which subsets of publications or filters can be projected" (Sjögårde, 2022c). In the first case study, I created overlays based on research publications addressing the Covid-19/SARS-CoV2 pandemic (hereafter coronavirus publications) as well as publications cited by this publication set. This case study shows how the maps can be used to explore research on a particular concept. In the second case study, I show how the maps can be used to compare the research orientation of organizations by creating overlays based on the publication output of three universities in Stockholm.

5.4.2 Results

I created a base map including 2.7 million PubMed articles from 2020-February 2022 (Figure 13). Going from left to right we find psychology, nursing, and public health in the bottom left. Clinical research is found in the upper left, whereas cell and molecular medicine are found in the top middle. Natural sciences are located in the middle right. Furthest to the right we find biochemistry and biophysics. Using the web-version[15] of this map, it is possible to zoom and navigate from the top-level nodes (disciplines) to their underlying specialties. Clicking a specialty highlights the most related specialties and a list of underlying topics is shown. Hyperlinks are also provided to the publications underlying each node.

[15] https://petersjogarde.github.io/papers/hiervis/base/index.html

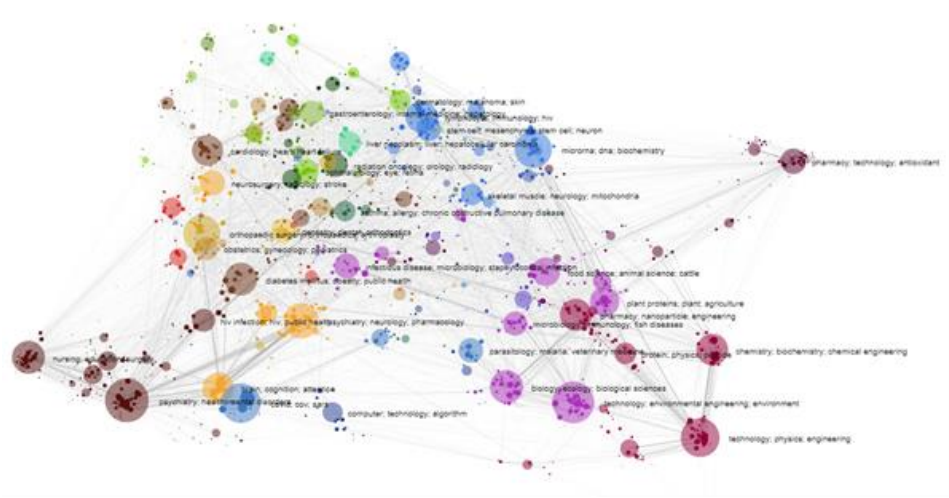

Figure 13: Base map of biomedical sciences[16] based on about 2.7 million articles from 2020-February 2022.

In the first case study, I created an overlay showing the distribution of coronavirus over clusters. Another overlay was created showing the distribution of citations from the coronavirus publications over clusters. This second map makes it possible to detect research fields of importance for the coronavirus research.

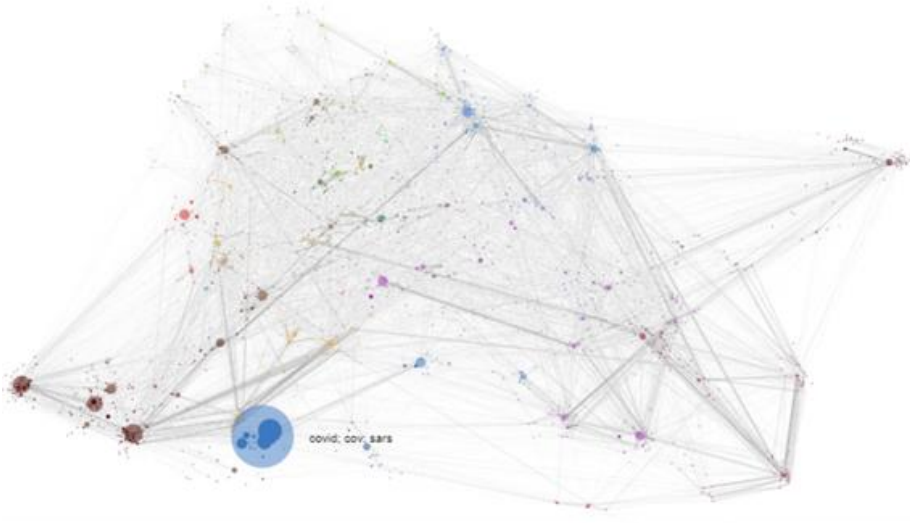

Figure 14: Covid-19/SARS-CoV-2 research[17].

[16] https://petersjogarde.github.io/papers/hiervis/base/index.html
[17] https://petersjogarde.github.io/papers/hiervis/covid_v2/pubs/index.html

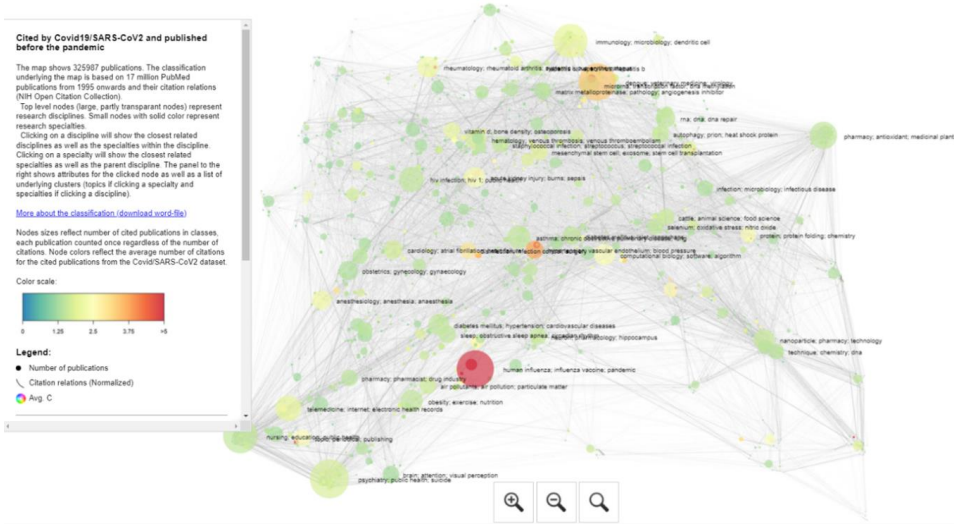

Figure 15: Publications cited by Covid-19/SARS-CoV-2 research[18], published before the pandemic. Node sizes reflect the number of cited publications and node colors the average number of citations per cited publication.

In the second case study, I show how the mapping methodology can be used to compare the subject orientation of universities. I created three overlays (Figure 16) based on the biomedical publications from (1) KTH Royal Institute of Technology (KTH), (2) Stockholm University (SU) and (3) Karolinska Institutet (KI). These three universities are part of an alliance called "Stockholm Trio". The maps show a focus on biochemistry and biophysics at KTH, an orientation towards natural sciences and psychology at SU and broad coverage of biomedicine at KI, including a wide range of clinical and basic research, but with few publications in natural sciences, biochemistry, and biophysics.

[18] https://petersjogarde.github.io/papers/hiervis/covid_v2/cited/index.html

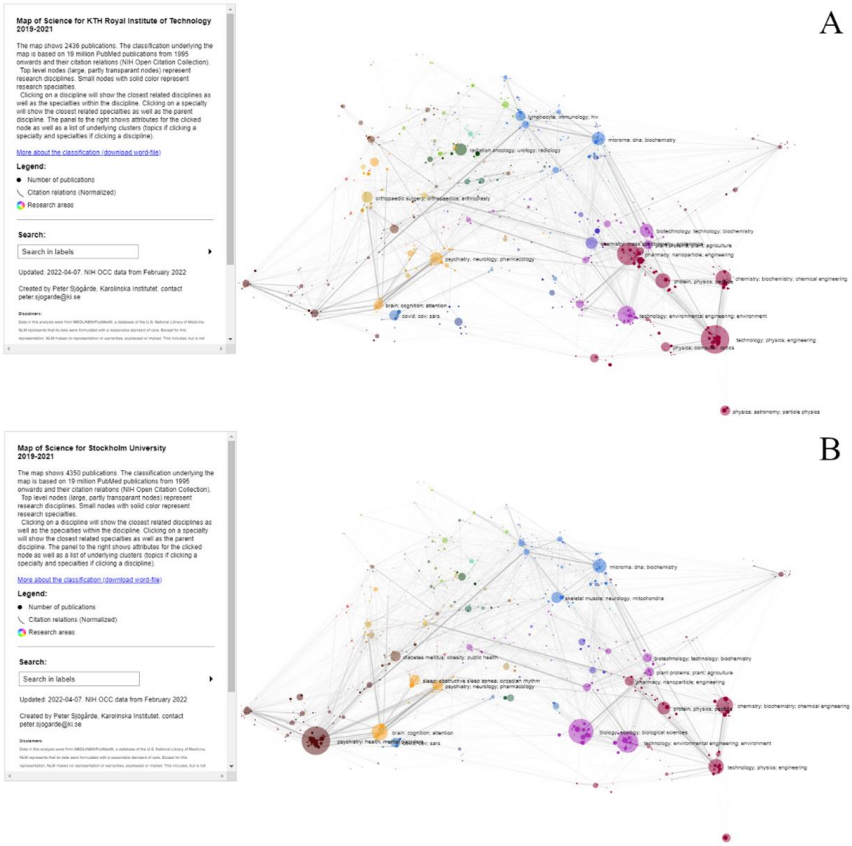

Figure 16: Biomedical publications by KTH[19] (A), Stockholm university[20] (B) and Karolinska institutet[21] (C). Publication years 2019-2021.

[19] https://petersjogarde.github.io/papers/hiervis/sthlm_trio/kth/index.html
[20] https://petersjogarde.github.io/papers/hiervis/sthlm_trio/sthlm_univ/index.html
[21] https://petersjogarde.github.io/papers/hiervis/sthlm_trio/ki/index.html

I conclude that the visualization methodology can be used to provide both overview and detail of large sets of research publications and that such maps "constitute a valuable tool for researchers studying science, improve transparency to cluster-based citation normalization, support research management and policy making and constitute a tool for researchers to explore research of relevance to them." (Sjögårde, 2022c)

As a result of Article IV, the classification was also made openly available[22].

[22] https://figshare.com/collections/PubMed_Classification/5610971

6 DISCUSSION

In this section, I first discuss the approach to use citations for the mapping of science (Section 6.1). I then discuss to what extent clusters represent research fields (6.2), followed by a discussion about the structural properties of research fields in relation to the properties of the obtained clusters (6.3). In Section 6.4, I discuss labeling and interpretation of clusters. Finally, I discuss applications of ACPLCs (6.5).

6.1 MAPPING SCIENCE THROUGH CITATIONS

Communication is essential within the science system and in the creation and formation of research fields. Citation networks offer a simplified representation of the communication taking place in the system. The focus on communication makes clustering of publications in citation networks a promising approach to obtain clusters corresponding to research fields. However, the simplification offered by direct citations reduces the relation between two publications to a binary entity. This means that the strength of the relation between a citing and cited publication is the same for any such pair of publications, regardless of the diverse meanings of citations, motivations to cite and varying relevance of cited publications to citing publications (Held & Velden, 2022). Furthermore, some citations may refer to publications focusing on other topics than the primary topic of the citing publication and relevant publications may be missing from the reference list, for example, publications that are unknown to the author. However, it is reasonable to assume, as done by the social systems citation theory, that the references in a publication have been selectively chosen by the author(s) based on their perceived relevance to some aspect of the publication (Tahamtan & Bornmann, 2022). Thereby, references contextualize and position each publication in the science system. Moreover, the clustering of a publication is affected by all of its relations to other publications, not by single edges to other publications. It is likely that this feature of the clustering methodology reduces issues associated with citation relations of less relevance. It has been shown that clustering approaches using larger sets of data ("global" models) result in both better precision and recall than approaches using less data ("local" models) and that the former approaches expand the context by including both publications in the research field of interest as well as publications in related research fields (Boyack, 2017; Klavans & Boyack, 2011, 2017b). The studies included in this thesis are based on large amounts of citation relations from long time periods, about 300 million to a billion citation relations covering more than 20 publication years. More than 90% of the publications in these networks have at least 5 direct citation relations. The scale of the studies can be assumed to offer some stability to the results and reduce problems associated with using binary relations between publications.

6.2 CLUSTERS AS REPRESENTATIONS OF RESEARCH FIELDS

Do clusters represent research fields, and if so, to what extent? In Article I we suggest that clusters can be used to identify topics if the granularity has been adjusted to obtain clusters corresponding to the size of topics. The rationale of the method used in Article I is that a

publication including a high number of references in general focuses on a topic and synthesizes the current state of the topic. This is a rough approximation, because the reference list in such publications may represent more than one topic and publications addressing the main topic of the publication may be missing. However, a perfect baseline of publications and their topics does not exist, and researchers perceive topics differently (Held et al., 2021). The results of Article I show that a rather high proportion of the references in the baseline papers are concentrated to one or a few clusters. The rest of the references are distributed over many clusters. This result corresponds well with the assumption that each synthesizing publication mainly focuses on one or just a few topics. Furthermore, the case studies indicate that the methodology could be used to identify topics. It was not difficult to interpret the topics represented in the clusters in the two case studies. However, it did take some effort to distinguish the topical orientation of a few of the clusters in one of the cases. This may indicate that some clusters do not coherently correspond to one topic.

In Article II, we created a baseline from a set of journals, arguing that a journal that is not very large or small in size, and has a focused scope, is likely to correspond to a specialty. This baseline focuses on communication, in that publication channels of specialties are used as proxies for publication classes corresponding to specialties. The restrictions we made reduce the risk of including journals having a broader or more narrow scope than specialties. Therefore, we assume that a majority of the journals in the baseline are focused to areas that, at least roughly, correspond to specialties. The results show that several granularity levels perform similarly, which indicates that the granularity of specialties is hard to determine. It may also suggest that the methodology does not capture specialties very well. It is possible that a better baseline can distinguish a granularity level corresponding to specialties more distinctly. Nevertheless, the case studies give some support for the correspondence between specialties and clusters, since specialties could be identified from the clusters and those specialties correspond rather well with other studies (Bauer et al., 2016; Blessinger & Frasier, n.d.; Figuerola et al., 2017; Janssens et al., 2006; Kim & Delen, 2018; Schuemie et al., 2009; L. Wang et al., 2017).

It is not always easy to distinguish between a discipline and a specialty because these distinctions may vary between fields and be more applicable in some fields and less suitable to describe other fields. Disciplines have not been in focus for this thesis. However, if disciplines are seen as broader focus areas of research than specialties, they should be more heterogenous and include different focus areas communicating in different communication channels and addressing different problem areas. The indicated definition seems to correspond rather well with, for example, the discipline library and information science and the specialties identified in Article II. The identified specialties have corresponding conferences and journals. Nevertheless, conferences and journals with both a broader and more narrow scope exist, which may suggest that the distinction between a discipline and a specialty is not clear-cut. Even though the existence of specialties is somewhat supported in this thesis, there is a lack of empirical evidence to support that this notion describes the structural properties of how research is self-organized into focus areas. It is possible that

different fields are structured in ways that are not accurately captured by the notion of specialties. This is indicated by the organizational structures of universities. There are large differences between for example, medicine, humanities, and social sciences in this regard, at least in Sweden. In medicine it is very common that the organizational structure is focused to research groups led by a professor. The department is secondary in these organizational structures. This is not the case in the social sciences and humanities, in which the department has a more important role for how research is organized, and groups may even be nonexistent. The size of departments may also vary tremendously, which may indicate that research fields have different structural properties. Another issue is multidisciplinary fields, which may gather researchers with different backgrounds. Researchers in such fields may study the same research topics but have weaker communication than suggested by the definition of specialties. ACPLCs may be one tool to better understand research fields, but there is a need to better understand how clusters correspond to structural properties in different fields.

Two case studies, one of "invasion biology" (Held & Velden, 2022) and another of "overall water splitting" (Haunschild et al., 2018), show that sets of publications constructed to represent research fields are distributed over a high number of clusters in ACPLCs and the authors suggest that ACPLCs do not properly delineate publications into research fields. These studies share a weakness: they both assume that the search queries they use accurately identify publications in the research fields of interest. However, problems with the search queries can easily be found. For example, both studies use the topic search tag ("TS") in Web of Science. This tag includes searching in "keywords plus". Keywords plus are assigned to publications based on titles of their references, not based on the bibliographic information of the publication itself. Hence, the search may capture publications related to the search terms merely by the titles in the reference lists. The search by Held and Velden (2022) also includes problematic terms, such as "non-indigenous community" or "exotic range" which are both terms used in other contexts than their field of interest (invasion biology).[23] Furthermore, it is widely known among search experts that it is extremely difficult to achieve search results with both high precision (the proportion of publications retrieved by the search query that are relevant) and high recall (the proportion of all relevant publications that are retrieved by the search query), for example, when conducting a systematic review or meta-analysis. One can assume that the same level of difficulty applies to the identification of publications belonging to a research field. It cannot be concluded from these studies that the clusters represent the research fields of interest with less accuracy than the set of publications retrieved by the search queries used for evaluations. Moreover, in both studies there is in fact a cluster identified that corresponds to the field of interest ("invasion biology" in Held and Velden

[23] "non-indigenous community" is used about the group of people not being indigenous in a geographic area and "exotic range" is used in physics.

(2022) and “overall water splitting” in Haunschild et al. (2018)). This may as well be interpreted as the main topic of the publication set could be identified.

This brings us to a challenging problem. How can and should clusters be evaluated? All approaches to evaluate clusters, that I am aware of, have weaknesses. The assessment of clusters by expert judgement is subject to bias and experts generally have a rather narrow perspective centered to their field of expertise and lack knowledge in classification. Baselines can be used to compare different methodological approaches but are less useful to evaluate to which extent clusters are meaningful or useful to users. The problem of evaluating classifications is not restricted to clusters obtained in citation networks. In fact, all classifications of research publications into research fields show at least some weaknesses when assessed. For example, Wang and Waltman (2016) found that the journal classification systems in Web of Science and Scopus assign journals to categories rather liberally, resulting in a rather low precision in the categories. Recent studies have shown large discrepancies between different approaches for the identification of publications addressing the sustainable development goals (Purnell, 2022; Rafols et al., 2021). Different studies on artificial intelligence (AI) result in very different operationalizations of the publications related to AI (e.g., see X. Chen et al., 2020; Di Vaio et al., 2020; Munim et al., 2020; Ruiz-Real et al., 2020; Sjögårde, 2022b). The precision of search queries used in systematic reviews is generally very low because systematic reviews favor recall (Bramer et al., 2016; Gusenbauer & Haddaway, 2020). These are all indications of how difficult it is to identify the publications related to a particular research field or, in the case of systematic reviews, the publications related to a particular research question. Given these difficulties, it is very problematic to construct sets of publications that can be used to assess cluster solutions. One can question if it is at all meaningful to evaluate clustering solutions using such approaches (Peel et al., 2017). A solution could be to evaluate the appropriateness and usefulness of ACPLCs and other classifications in relation to intended use. However, there is a lack of such studies for the assessment of clustering solutions. A couple of exceptions are the studies by Perianes-Rodriguez and Ruiz-Castillo (2017) and Ruiz-Castillo and Waltman (2015) addressing the suitability of different classifications for normalization of citation counts. Another exception is a small study conducted by me focusing on information needs of managers in a project promoting competence development in artificial intelligence and how ACPLC could be used to correspond to the identified needs (Sjögårde, 2022b). The study showed a need from the management of the project for more information about who was doing what within AI and to get an overview of the research applying AI. This study is too small to draw any general conclusions. However, the study provides an example of how ACPLCs can be used to support research management and development of support structures.

6.3 STRUCTURAL PROPERTIES OF RESEARCH FIELDS AND CLUSTERS

In the entry for “classification” in the International Society of Knowledge Organization (ISKO) Encyclopedia of Knowledge Organization Hjørland distinguishes between two views on classification, the “classical view” and the “prototype theory” (Hjørland, 2020). In the

classical view on classification, originating from Aristotle, “classes should be designed so membership of a class is given by a set of necessary and sufficient characteristics” (Hjørland, 2020). All items of a class must fulfill such a criterion and, hence, all items of a class share one characteristic or a set of characteristics. The relation between an item and a class is binary in the sense that the item either does or does not fulfill the criterion.

The classical view is contrasted by the prototype theory. In this view, each class is characterized by a set of characteristics. The relation between an item and a class is determined by the correspondence of the publication to a prototype, an item possessing all of the characteristics being typical for the class. In this view, items may to varying extent be related to a class and “[d]issimilarity plays as important a role as similarity in classification. Similarity alone is not enough (see Andersen et al. 2006, 24ff).” (Hjørland, 2020). Hence, classes may include items that are more typical or central, and others that are less typical, or more peripheral. The borders of classes are by the prototype theory fuzzy since some of the members of a class may have a weak relation to the class. Furthermore, all members of a class do not need to share one characteristic or a set of characteristics and one pair of items in a class may share some characteristics and another pair of items in the same class another set of characteristics.

The prototype theory of classification provides an interesting framework for how to understand research fields. Research fields have a core-periphery structure in that some research is a better representative of the field and some research may also influence the field to a higher extent. Furthermore, research fields have fuzzy borders and are overlapping. The prototype theory captures such complex structures better than the classical view. If we consider these properties in relation to the classification of research publications into research fields, we see that the prototype theory allows (1) some publications to be better representatives of a research fields than others, (2) publications to be partly related to a research field, (3) fuzzy borders of publication classes, and (4) overlap of publication classes. Some of these properties may be allowed by the classical view as well, but not all and not to the same extent.

The clusters obtained by the Leiden algorithm do not overlap, i.e., each publication belongs to exactly one cluster. Furthermore, publications in the obtained classifications have binary relations to clusters, i.e., a publication either belongs to a cluster or it does not. These properties are problematic if clusters are to represent research fields because it neither allow research fields to overlap nor for some publications to have stronger relations to a cluster than others. Other algorithms may capture overlapping structures in the citation network, such as the OSLOM algorithm (Lancichinetti et al., 2011) or the IKC algorithm combined with AOC (Jakatdar et al., 2022; Wedell et al., 2022). Nevertheless, neither of these approaches express varying relational strengths between items and clusters.

If we consider an ACPLC as part of a larger model, it also includes the citation network on which the clusters have been based. Taking this more holistic view, the model includes properties of overlap and the possibility to calculate strength of publication-cluster relations.

Citation relations between clusters can be used to indicate relations between research fields. Publications having strong citation relations to external clusters are likely related to more than one research field and may be used to express overlap of topics. This holistic view offers the possibility to use and develop ACPLCs in a way that overcomes the inconsistency between the definition of topics and specialties given in this thesis and the disjoint clusters, as well as the binary relation between clusters and publications. However, further work is needed in this area.

Figure 17 shows a rather typical core-periphery structure that appears when clusters at the topic level are visualized. Nodes represent publications, edges represent normalized citation relations, and node sizes represent the number of citation relations a publication has within the cluster. Core publications are positioned in the center of the graph and have many relations, while peripheral publications are located in the marginals and have few relations. The topic to which the cluster corresponds (in this case "peptic ulcer perforation") is likely to be better represented in the core of the cluster.

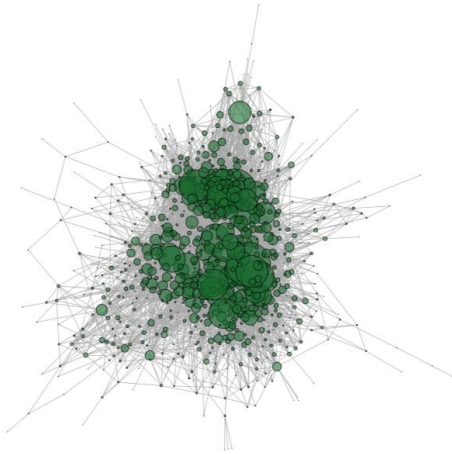

Figure 17: Illustration of the core-periphery structure of a cluster.[24] *Nodes represent publications and edges represent normalized direct citation relations. Node sizes correspond to citation relations within the cluster.*

6.4 LABELING AND INTERPRETATION OF CLUSTERS

A meaningful interpretation of a cluster can only be done if the publications in the cluster have been grouped meaningfully. To be meaningful, a cluster should first of all make intuitive sense, at least to subject experts (Šubelj et al., 2016). To make intuitive sense, a cluster should be concentrated to a focus area in a somewhat coherent way, meaning that at least the publications in the core of a cluster should be semantically related, for example by

[24] The illustrated cluster (id=4898) is from the September 2022 update of the PubMed classification available in Figshare: https://figshare.com/articles/dataset/PubMed_classification_v1_202102/16601402/1

focusing on an empirical object, a particular process, a set of methods, or a combination thereof. Furthermore, it should be possible to distinguish a cluster from other clusters. Also, the clustering of sub-level clusters into parent clusters should make intuitive sense. If the clusters do not fulfill these criteria of meaningfulness, at least to an acceptable extent, it is not expected that they can be meaningfully labeled. However, if clusters do fulfill such criteria, labeling is possible and should provide the possibility, at least for subject experts, to interpret the focus area represented in a cluster and to differentiate this focus area from the focus areas of other clusters.

The labeling approach proposed in Article III relies on the assumption that clusters are meaningful. We argue that terms that are both frequent in a cluster and specific to the cluster, in comparison to sibling clusters, are likely to express the focus area of the cluster. In practice, not all clusters include terms that are both frequent in the cluster and specific to the cluster. Instead, there is often a tradeoff between frequency and specificity. There is a variation between clusters regarding how the highest ranked terms in the clusters are balanced by the TFS approach. Some clusters at the same granularity level are labeled by terms with high frequency, while others are labeled with terms with high specificity. This is problematic because it makes interpretation more difficult.

Another challenge is that labels may be easy to interpret but inaccurate in the sense that they correspond poorly to the focus area of the publications in the cluster. This problem does not only apply to clusters, but also other kinds of categorizations of research publications, such as journal categories, search results or other sets of publications used to delineate publications into fields. To illustrate the problem, we may use an example. Assume that a user is provided with the information that there are 13,000 publications in a cluster labeled “alzheimer disease”. The focus area can be easily interpreted from the label. Nevertheless, the label may correspond poorly to the focus area of the publications in the cluster, and it is very hard for users to judge whether the label accurately represents the set of 13,000 publications, or not. An example of when a label may correspond poorly to the focus area of the publications in a cluster is when a cluster has been labeled by terms that merely represent a subfield within the cluster. In the Alzheimer example, the cluster may for instance be focusing on other types of dementia as well.

Experience from the studies in this thesis, as well as from using clustering in practice, indicates that interpreting the subject orientation of clusters is sometimes challenging (Sjögårde, 2022b). Interpretation is supported by the possibility to navigate the ACPLC and get information about the relation with other clusters, such as parent cluster, sibling clusters and underlying clusters. This information makes it easier to distinguish the subject orientation of one cluster from other clusters. Providing users with information about frequency and specificity of the labels may further facilitate interpretation, because it makes it possible to judge how well the label represents the publications of the cluster.

6.5 APPLICATIONS OF ACPLCS

A perfect delineation of publications into research fields is not feasible given the prototype theory of classification and the properties of research fields that I have outlined, including a core-periphery structure and fuzzy borders. It is difficult to see how the boundaries of a research field can be determined without some degree of arbitrariness. Therefore, we cannot expect that the publications related to a research field can be delineated unambiguously. Furthermore, it is not known how well clusters correspond to research fields. For example, it might be the case that some clusters at the topic level correspond to more than one topic, and that some topics are represented by several clusters. This brings us to the question about when it is appropriate to use an ACPLCs.

To begin with, ACPLCs are useful to *identify* and *explore* focus areas of research within sets of publications, in particular large sets of publications. ACPLCs offer a possibility to get overviews of sets of publications that are far larger than what can be browsed and processed manually. This is illustrated in the two case studies in Article IV, in which overviews are given of corona virus research as well as the biomedical research of three universities in Stockholm. The use of a hierarchical ACPLC makes it possible to provide both overview and detail, which would not have been possible using a one-level classification, such as for example the commonly used Web of Science journal classification. Related to this application is the use of ACPLCs to display and explore search results (Bascur et al., 2019, 2020). In addition, clusters can be used both to restrict search results and to broaden search results.

A second area of application is the use of ACPLCs to establish general patterns or relationships. For such use it may be acceptable with an approximate delineation into research fields. For example, in Ahlgren et al. (2018) we studied the relation between citation counts and properties of references in publications. We used fine-grained clusters to normalize citation counts with the purpose to control for field differences. Another example is in Sjögårde and Didegah (2022) where we studied the relation between citation counts and growth of research fields. We showed that publications in growing fields generally have a citation advantage because of the increasing number of publications citing earlier publications within the field.

A third application is the use of ACPLCs as data input in other calculations, for example to measure interdisciplinarity (Q. Wang & Ahlgren, 2018), or to identify benchmark units (Q. Wang & Jeppsson, 2022). Another example is an application that I have elaborated together with colleagues at the KI library, in which we used clusters as variables in a supervised classification approach that assigns publications to classes in a national classification system of research subjects.[25] Except for clusters we used MeSH and classifications of researchers to

[25] Standard för svensk indelning av forskningsämnen 2011. URL: https://www.scb.se/dokumentation/klassifikationer-och-standarder/standard-for-svensk-indelning-av-forskningsamnen/

train a neural network. When adding cluster information, the accuracy of the classification increased.

A fourth application is the use of ACPLCs for normalization of citation indicators. There is some support that clusters are preferred for citation normalization over the frequently used Web of Science journal classification (Perianes-Rodriguez & Ruiz-Castillo, 2017; Ruiz-Castillo & Waltman, 2015). However, other approaches to normalize citations exist, such as source normalization and item-oriented approaches (Colliander, 2015; Colliander & Ahlgren, 2019; Waltman & van Eck, 2013a).

It is also interesting to ask in what situations ACPLCs should not be utilized. In my experience, it seems that there is seldom a one-to-one relation between a cluster in an ACPLC and a predefined research field. This is confirmed in the studies of Haunschild et al. (2018) and Held and Velden (2022). Since the clusters are not overlapping, it is unlikely that clusters offer a high recall of the publications in a research field. This problem may be tackled by approaches providing overlap by, for example, expanding clusters. However, such approaches may result in lower precision.

7 CONCLUSIONS

In this thesis, I have identified research fields in the science system through the network structures created by researchers when contextualizing their research by the use of references in their publications. The methodology that I have applied provides overview and possibilities to explore very large amounts of data. I have elaborated a conceptual framework and adjusted the granularity of clusters in order to obtain clusters corresponding to the size of topics and specialties, which I have defined as focus areas of research at different granularity levels. I argue that clusters at the most granular level obtained in the thesis generally represent topics and at the next level specialties.

Labels improve the utility of ACPLCs by making it possible to interpret the research fields represented by clusters. However, labeling is a challenging problem, not the least because of the different semantic levels of focus areas in clusters of different granularities. In Article III, we suggest that labels can be assigned to clusters using various bibliographic information combined with a term weighting approach that is able to assign broad, general terms at higher semantic levels and narrow, specific terms at lower semantic levels. Thereby, interpretability of ACPLCs can be facilitated. Moreover, I have proposed a methodology to visualize ACPLCs that further improves interpretation by making it possible to explore the context of a cluster, for instance by displaying information about the parent cluster, underlying clusters, and related clusters. Interpretability and exploration are also improved by hyperlinks to the publications included in the clusters.

Visualization is needed in some applications of ACPLCs. Previous visualizations of classifications, for example overlay maps based on journal categories, have not made it possible to get both overview and detail of the science system. I show in Article IV how this can be achieved by visualizing multiple levels in an ACPLC and by making the visualization interactive and possible to navigate.

There are still several ways in which ACPLC can be further improved. First of all, the core-periphery structure of research fields and the overlapping nature of fields are not expressed in ACPLCs. Future work may address how such structures can be incorporated in science mapping. A second issue is the binary use of citations, which may neglect important information. Future work may explore how “epistemic functions” can be used to develop ACPLCs representing research fields in richer, more multidimensional manners (Held & Velden, 2022). A third issue for future work is to further improve cluster labeling. The labeling procedure elaborated in this thesis captures relevant terms at different granularity levels, but does not consider semantic relations between terms. Future work could address this problem by developing a labeling approach that considers semantic relations, for example by making use of language models such as SciBERT (Beltagy et al., 2019). A fourth research theme for future work is the evaluation of ACPLCs in relation to different use cases. Such research may establish more knowledge about how ACPLCs are interpreted by users and find out to which user needs ACPLCs contribute with valuable information.

Critique has been raised that ACPLCs are not transparent. I have shown that this problem is possible to overcome by providing labels and additional information to clusters, using well documented and openly available methods, and visualizing ACPLCs so they become explorable. In addition, I have based the ACPLC in Article IV on open data and published the ACPLC openly available. In fact, this transparency goes beyond the transparency of most other classifications used within the field of scientometrics.

The aim of this thesis has been to improve the interpretability and utility of ACPLCs. I have met this aim by providing a conceptual framework, adjusted granularity so that the size of clusters corresponds to topics and specialties, suggested a method to provide labels at different semantic levels and visualized the ACPLC in a way that provides both overview and detail. As a result of this work, I have created a classification of about 20 million PubMed records[26] that has been made openly available (Sjögårde, 2022a). I hope others find the classification useful.

[26] https://figshare.com/collections/PubMed_Classification/5610971

8 ACKNOWLEDGEMENTS

I would like to thank all of you who have been giving me support during this project. I would like to express my gratitude in particular to:

Min fru Linnéa för den kärlek och det stöd du alltid ger mig.

Mina söner, Boe och Vigge, som båda föddes under arbetet med denna avhandling. Jag vill tacka er för de frågor ni ställer och för det nyfikna, orädda, slumpartade och galna sätt som ni utforskar världen med. Ni är förebilder för vetenskapen!

Min familj och mina vänner för ert stöd och uppmuntran.

All of you involved in realizing this project at Karolinska Institutet, Erik Stattin, Annikki Roos, Carl Johan Sundberg, Sabine Koch and Catharina Rehn. This project would not have been realized without your efforts and support.

The Foundation for Promotion and Development of Research at Karolinska Institutet for funding this project.

My supervisors, Sabine Koch, Carl Johan Sundberg, Per Ahlgren and Ludo Waltman, for your support and for generously sharing your knowledge and taking time to comment and discuss my work.

All my colleagues at the university library for your support, discussions, and for sharing your knowledge and expertise. I would like to give a special thanks to the bibliometric analyst group, Catharina Rehn, Robert Juhasz, Fereshteh Didegah and Alvin Gavel.

The Health Informatics Group at the Department of LIME and all others at LIME who have been giving feedback and support.

My former colleagues at the Unit for Publication Infrastructure, for providing the inspiring environment where I started out working with clustering-based approaches.

Lennart Stenberg at Vinnova for your enthusiasm which contributed to setting me on this path.

The Department of ALM in Uppsala where I took the first steps of this journey.

Everyone who has been taking their time to read and comment on my work, including half-time committee, Jussi Karlgren, Martin Rosvall and Wolfgang Glänzel, pre-dissertation seminar discussant, Vincent Traag, and all others asking questions, giving comments, and discussing my work at seminars, conferences and in papers.

9 REFERENCES

Abramo, G., D'Angelo, C. A., & Zhang, L. (2018). A comparison of two approaches for measuring interdisciplinary research output: The disciplinary diversity of authors vs the disciplinary diversity of the reference list. *Journal of Informetrics*, *12*(4), 1182–1193. https://doi.org/10.1016/j.joi.2018.09.001

Ahlgren, P., Chen, Y., Colliander, C., & van Eck, N. J. (2020). Enhancing direct citations: A comparison of relatedness measures for community detection in a large set of PubMed publications. *Quantitative Science Studies*, *1*(2), 1–17. https://doi.org/10.1162/qss_a_00027

Ahlgren, P., & Colliander, C. (2009a). Textual content, cited references, similarity order, and clustering: An experimental study in the context of science mapping. In B. Larsen & J. Leta (Eds.), *Proceedings of Issi 2009—12th International Conference of the International Society for Scientometrics and Informetrics, Vol 2* (Vol. 2, pp. 862–873). Int Soc Scientometrics & Informetrics-Issi.

Ahlgren, P., & Colliander, C. (2009b). Document–document similarity approaches and science mapping: Experimental comparison of five approaches. *Journal of Informetrics*, *3*(1), 49–63. https://doi.org/10.1016/j.joi.2008.11.003

Ahlgren, P., Colliander, C., & Sjögårde, P. (2018). Exploring the relation between referencing practices and citation impact: A large-scale study based on Web of Science data. *Journal of the Association for Information Science and Technology*, *69*(5), 728–743. https://doi.org/10.1002/asi.23986

Ahlgren, P., & Jarneving, B. (2008). Bibliographic coupling, common abstract stems and clustering: A comparison of two document-document similarity approaches in the context of science mapping. *Scientometrics*, *76*(2), 273–290. https://doi.org/10.1007/s11192-007-1935-1

Ahlgren, P., Jarneving, B., & Rousseau, R. (2003). Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient. *Journal of the American Society for Information Science and Technology*, *54*(6), 550–560. https://doi.org/10.1002/asi.10242

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open*, *9*(1), 2158244019829575. https://doi.org/10.1177/2158244019829575

Amsterdamska, O., & Leydesdorff, L. (1989). Citations: Indicators of significance? *Scientometrics*, *15*(5), 449–471. https://doi.org/10.1007/BF02017065

Bascur, J. P., van Eck, N. J., & Waltman, L. (2019). An interactive visual tool for scientific literature search: Proposal and algorithmic specification. *BIR 2019 Workshop on Bibliometric-Enhanced Information Retrieval*, *2414*, 76–87.

Bascur, J. P., Verberne, S., van Eck, N. J., & Waltman, L. (2020). Browsing citation clusters for academic literature search: A simulation study with systematic reviews. *Proceedings of the 10th International Workshop on Bibliometric-Enhanced Information Retrieval*, *2591*, 53–65.

Bassecoulard, E., & Zitt, M. (1999). Indicators in a research institute: A multi-level classification of scientific journals. *Scientometrics*, *44*(3), 323–345. https://doi.org/10.1007/BF02458483

Bauer, J., Leydesdorff, L., & Bornmann, L. (2016). Highly cited papers in Library and Information Science (LIS): Authors, institutions, and network structures. *Journal of the*

Association for Information Science and Technology, *67*(12), 3095–3100. https://doi.org/10.1002/asi.23568

Beltagy, I., Lo, K., & Cohan, A. (2019). *SciBERT: A pretrained language model for scientific text* (arXiv:1903.10676). arXiv. https://doi.org/10.48550/arXiv.1903.10676

Blessinger, K., & Frasier, M. (n.d.). Analysis of a decade in library literature: 1994–2004. *College & Research Libraries*, *68*(2), 155–170. https://doi.org/10.5860/crl.68.2.155

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Börner, K., Chen, C., & Boyack, K. W. (2005). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, *37*(1), 179–255. https://doi.org/10.1002/aris.1440370106

Bornmann, L., Marx, W., & Barth, A. (2013). The normalization of citation counts based on classification systems. *Publications*, *1*(2), 78–86. https://doi.org/10.3390/publications1020078

Boyack, K. W. (2017). Investigating the effect of global data on topic detection. *Scientometrics*, *111*(2), 999–1015. https://doi.org/10.1007/s11192-017-2297-y

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, *61*(12), 2389–2404. https://doi.org/10.1002/asi.21419

Boyack, K. W., & Klavans, R. (2014). Including cited non-source items in a large-scale map of science: What difference does it make? *Journal of Informetrics*, *8*(3), 569–580. https://doi.org/10.1016/j.joi.2014.04.001

Boyack, K. W., & Klavans, R. (2018). Accurately identifying topics using text: Mapping PubMed. *STI 2018 Conference Proceedings*, *23*. https://openaccess.leidenuniv.nl/handle/1887/65319

Boyack, K. W., & Klavans, R. (2020). A comparison of large-scale science models based on textual, direct citation and hybrid relatedness. *Quantitative Science Studies*, 1–16. https://doi.org/10.1162/qss_a_00085

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, *64*(3), 351–374. https://doi.org/10.1007/s11192-005-0255-6

Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., Schijvenaars, B., Skupin, A., Ma, N., & Börner, K. (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. *PLOS ONE*, *6*(3), e18029. https://doi.org/10.1371/journal.pone.0018029

Braam, R. R., Moed, H. F., & Raan, A. F. J. van. (1991). Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for Information Science*, *42*(4), 233–251. https://doi.org/10.1002/(SICI)1097-4571(199105)42:4<233::AID-ASI1>3.0.CO;2-I

Bramer, W. M., Giustini, D., & Kramer, B. M. R. (2016). Comparing the coverage, recall, and precision of searches for 120 systematic reviews in Embase, MEDLINE, and Google Scholar: A prospective study. *Systematic Reviews*, *5*(1), 39. https://doi.org/10.1186/s13643-016-0215-7

Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, *22*(1), 155–205. https://doi.org/10.1007/BF02019280

Callon, M., Courtial, J.-P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, *22*(2), 191–235. https://doi.org/10.1177/053901883022002003

Cao, M. D., & Gao, X. (2005). Combining contents and citations for scientific document classification. In S. Zhang & R. Jarvis (Eds.), *AI 2005: Advances in Artificial Intelligence* (pp. 143–152). Springer. https://doi.org/10.1007/11589990_17

Carmel, D., Roitman, H., & Zwerdling, N. (2009). Enhancing cluster labeling using Wikipedia. *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 139–146. https://doi.org/10.1145/1571941.1571967

Carpenter, M. P., & Narin, F. (1973). Clustering of scientific journals. *Journal of the American Society for Information Science*, *24*(6), 425–436. https://doi.org/10.1002/asi.4630240604

Chen, C.-M. (2008). Classification of scientific networks using aggregated journal-journal citation relations in the Journal Citation Reports. *Journal of the American Society for Information Science and Technology*, *59*(14), 2296–2304. https://doi.org/10.1002/asi.20935

Chen, P., & Redner, S. (2010). Community structure of the physical review citation network. *Journal of Informetrics*, *4*(3), 278–290. https://doi.org/10.1016/j.joi.2010.01.001

Chen, X., Chen, J., Cheng, G., & Gong, T. (2020). Topics and trends in artificial intelligence assisted human brain research. *PLOS ONE*, *15*(4), e0231192. https://doi.org/10.1371/journal.pone.0231192

Chubin, D. E. (1976). State of the field: The conceptualization of scientific specialties. *Sociological Quarterly*, *17*(4), 448–476. https://doi.org/10.1111/j.1533-8525.1976.tb01715.x

Colliander, C. (2014). *Science mapping and research evaluation: A novel methodology for creating normalized citation indicators and estimating their stability* [Doctoral thesis]. http://www.diva-portal.org/smash/record.jsf?pid=diva2:752675

Colliander, C. (2015). A novel approach to citation normalization: A similarity-based method for creating reference sets. *Journal of the Association for Information Science and Technology*, *66*(3), 489–500. https://doi.org/10.1002/asi.23193

Colliander, C., & Ahlgren, P. (2019). Comparison of publication-level approaches to ex-post citation normalization. *Scientometrics*, *120*(1), 283–300. https://doi.org/10.1007/s11192-019-03121-z

Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. University Of Chicago Press.

Cronin, B. (1984). *The citation process: The role and significance of citations in scientific communication*. T. Graham.

Davidson, G. S., Wylie, B. N., & Boyack, K. W. (2001). Cluster stability and the use of noise in interpretation of clustering. *Proceedings of the IEEE Symposium on Information Visualization 2001 (INFOVIS'01)*, 23.

de Solla Price, D. J. (1965a). *Little science, big science*. Columbia U.P.

de Solla Price, D. J. (1965b). Networks of scientific papers. *Science*, *149*(3683), 510–515. https://doi.org/10.1126/science.149.3683.510

de Solla Price, D. J., & Beaver, D. (1966). Collaboration in an invisible college. *American Psychologist*, *21*, 1011–1018. https://doi.org/10.1037/h0024051

Di Vaio, A., Palladino, R., Hassan, R., & Escobar, O. (2020). Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. *Journal of Business Research*, *121*, 283–314. https://doi.org/10.1016/j.jbusres.2020.08.019

Doyle, L. B. (1962). Indexing and abstracting by association. *American Documentation*, *13*(4), 378–390. https://doi.org/10.1002/asi.5090130404

Engerer, V. (2017). Exploring interdisciplinary relationships between linguistics and information retrieval from the 1960s to today. *Journal of the Association for Information Science and Technology*, *68*(3), 660–680. https://doi.org/10.1002/asi.23684

Fano, R. M. (1956). Information theory and the retrieval of recorded information. In J. H. Shera, A. Kent, & J. W. Perry, *Documentation in action*. Reinhold publishing corporation.

Figuerola, C. G., García Marco, F. J., & Pinto, M. (2017). Mapping the evolution of library and information science (1978–2014) using topic modeling on LISA. *Scientometrics*, *112*(3), 1507–1535. https://doi.org/10.1007/s11192-017-2432-9

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3–5), 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

Fortunato, S., & Barthélemy, M. (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences*, *104*(1), 36–41. https://doi.org/10.1073/pnas.0605965104

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*(11), 1129–1164. https://doi.org/10.1002/spe.4380211102

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, *122*(3159), 108–111. https://doi.org/10.1126/science.122.3159.108

Garfield, E. (1963). Citation indexes in sociological and historical research. *American Documentation*, *14*(4), 289–291. https://doi.org/10.1002/asi.5090140405

Garfield, E. (1964). Can citation indexing be automated? In *Essays of an information scientist* (Vol. 1, pp. 84–90).

Garfield, E. (1972). Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, *178*(4060), 471.

Garfield, E., Sher, I., & Torpie, R. J. (1964). *The use of citation data in writing the history of science*. https://doi.org/10.21236/ad0466578

Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science Advances*, *4*(7), eaaq1360. https://doi.org/10.1126/sciadv.aaq1360

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. https://doi.org/10.1007/BF02093621

Glänzel, W., & Schubert, A. (2003a). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, *56*(3), 357–367. https://doi.org/10.1023/A:1022378804087

Glänzel, W., & Schubert, A. (2003b). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, *56*(3), 357–367. https://doi.org/10.1023/A:1022378804087

Gläser, J., Glänzel, W., & Scharnhorst, A. (2017). Same data—different results? Towards a comparative approach to the identification of thematic structures in science. *Scientometrics*, *111*(2), 981–998. https://doi.org/10.1007/s11192-017-2296-z

Glenisson, P., Glänzel, W., Janssens, F., & De Moor, B. (2005). Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing & Management*, *41*(6), 1548–1572. https://doi.org/10.1016/j.ipm.2005.03.021

Gusenbauer, M. (2019). Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. *Scientometrics*, *118*(1), 177–214. https://doi.org/10.1007/s11192-018-2958-5

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, *11*(2), 181–217. https://doi.org/10.1002/jrsm.1378

Hagstrom, W. (1970). Factors related to the use of different modes of publishing research in four scientific fields. In E. C. Nelson & K. D. Pollock, *Communication among scientiests and engineers* (pp. 85–124). Heath Lexington Books.

Hammarfelt, B. (2019). Discipline. In *ISKO Encyclopedia of Knowledge Organization*. https://www.isko.org/cyclo/discipline

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. JSTOR. https://doi.org/10.2307/2346830

Haunschild, R., Schier, H., Marx, W., & Bornmann, L. (2018). Algorithmically generated subject categories based on citation relations: An empirical micro study using papers on overall water splitting. *Journal of Informetrics*, *12*(2), 436–447. https://doi.org/10.1016/j.joi.2018.03.004

Havemann, F., Gläser, J., & Heinz, M. (2017). Memetic search for overlapping topics based on a local evaluation of link communities. *Scientometrics*, *111*(2), 1089–1118. https://doi.org/10.1007/s11192-017-2302-5

Held, M., & Velden, T. (2022). How to interpret algorithmically constructed topical structures of scientific fields? A case study of citation-based mappings of the research specialty of invasion biology. *Quantitative Science Studies*, *3*(3), 651–671. https://doi.org/10.1162/qss_a_00194

Hjørland, B. (1992). The concept of 'subject' in information science. *Journal of Documentation*, *48*(2), 172–200. https://doi.org/10.1108/eb026895

Hjørland, B. (2002). Domain analysis in information science—Eleven approaches—Traditional as well as innovative. *Journal of Documentation*, *58*(4), 422–462. https://doi.org/10.1108/00220410210431136

Hjørland, B. (2020). Classification. In *ISKO Encyclopedia of Knowledge Organization*. https://www.isko.org/cyclo/classification

Hutchins, B. I. (2021). A tipping point for open citation data. *Quantitative Science Studies*, *2*(2), 433–437. https://doi.org/10.1162/qss_c_00138

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, *9*(6), e98679. https://doi.org/10.1371/journal.pone.0098679

Jakatdar, A., Liu, B., Warnow, T., & Chacko, G. (2022). AOC: Assembling overlapping communities. *Quantitative Science Studies*, *3*(4), 1079–1096. https://doi.org/10.1162/qss_a_00227

Janssens, F., Glänzel, W., & De Moor, B. (2008). A hybrid mapping of information science. *Scientometrics*, *75*(3), 607–631. https://doi.org/10.1007/s11192-007-2002-7

Janssens, F., Leta, J., Glänzel, W., & De Moor, B. (2006). Towards mapping library and information science. *Information Processing & Management*, *42*(6), 1614–1642. https://doi.org/10.1016/j.ipm.2006.03.025

Janssens, F., Zhang, L., Moor, B. D., & Glänzel, W. (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing & Management*, *45*(6), 683–702. https://doi.org/10.1016/j.ipm.2009.06.003

Jarneving, B. (2007). Bibliographic coupling and its application to research-front and other core documents. *Journal of Informetrics*, *1*(4), 287–307. https://doi.org/10.1016/j.joi.2007.07.004

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, *31*(1), 7–15. https://doi.org/10.1016/0020-0190(89)90102-6

Katz, J. S., & Hicks, D. (1995). The Classification of Interdisciplinary Journals: A New Approach. *Proceeding of The Fifth Biennial Conference of The International Society for Scientometrics and Informatics*, 105–115.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1), 10–25. https://doi.org/10.1002/asi.5090140103

Kessler, M. M. (1965). Comparison of the results of bibliographic coupling and analytic subject indexing. *American Documentation*, *16*(3), 223–233. https://doi.org/10.1002/asi.5090160309

Kim, Y.-M., & Delen, D. (2018). Medical informatics research trend analysis: A text mining approach. *Health Informatics Journal*, *24*(4), 432–452. https://doi.org/10.1177/1460458216678443

Klavans, R., & Boyack, K. W. (2006). Quantitative evaluation of large maps of science. *Scientometrics*, *68*(3), 475–499. https://doi.org/10.1007/s11192-006-0125-x

Klavans, R., & Boyack, K. W. (2011). Using global mapping to create more accurate document-level maps of research fields. *Journal of the American Society for Information Science and Technology*, *62*(1), 1–18. https://doi.org/10.1002/asi.21444

Klavans, R., & Boyack, K. W. (2017a). Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? *Journal of the Association for Information Science and Technology*, *68*(4), 984–998. https://doi.org/10.1002/asi.23734

Klavans, R., & Boyack, K. W. (2017b). Research portfolio analysis and topic prominence. *Journal of Informetrics*, *11*(4), 1158–1174. https://doi.org/10.1016/j.joi.2017.10.002

Koopman, R., Wang, S., & Scharnhorst, A. (2017). Contextualization of topics: Browsing through the universe of bibliographic information. *Scientometrics*, *111*(2), 1119–1139. https://doi.org/10.1007/s11192-017-2303-4

Kuhn, T. S. (1996). *The structure of scientific revolutions* (3rd edition). University of Chicago Press.

Lambiotte, R., & Panzarasa, P. (2009). Communities, knowledge creation, and information diffusion. *Journal of Informetrics*, *3*(3), 180–190. https://doi.org/10.1016/j.joi.2009.03.007

Lancichinetti, A., Radicchi, F., Ramasco, J. J., & Fortunato, S. (2011). Finding Statistically Significant Communities in Networks. *PLOS ONE*, *6*(4), e18961. https://doi.org/10.1371/journal.pone.0018961

Leydesdorff, L. (1987). Various methods for the mapping of science. *Scientometrics*, *11*(5–6), 295–324. https://doi.org/10.1007/BF02279351

Leydesdorff, L. (2004). Clusters and maps of science journals based on bi-connected graphs in the journal citation reports. *J. Documentation*. https://doi.org/10.1108/00220410410548144

Leydesdorff, L. (2006). Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? *Journal of the American Society for Information Science and Technology*, *57*(5), 601–613. https://doi.org/10.1002/asi.20322

Leydesdorff, L., Zhou, P., & Bornmann, L. (2013). How can journal impact factors be normalized across fields of science? An assessment in terms of percentile ranks and fractional counts. *Journal of the American Society for Information Science and Technology*, *64*(1), 96–107. https://doi.org/10.1002/asi.22765

Lin, J., & Wilbur, W. J. (2007). PubMed related articles: A probabilistic topic-based model for content similarity. *BMC Bioinformatics*, *8*, 423. https://doi.org/10.1186/1471-2105-8-423

Luukkonen, T. (1997). Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis. *Scientometrics*, *38*(1), 27–37. https://doi.org/10.1007/BF02461121

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. https://doi.org/10.1002/(SICI)1097-4571(198909)40:5<342::AID-ASI7>3.0.CO;2-U

Mai, J. (2011). The modernity of classification. *Journal of Documentation*, *67*(4), 710–730. https://doi.org/10.1108/00220411111145061

Mao, X.-L., Ming, Z.-Y., Zha, Z.-J., Chua, T.-S., Yan, H., & Li, X. (2012). Automatic labeling hierarchical topics. *Proceedings of the 21st ACM International Conference on Information and Knowledge Management - CIKM '12*, 2383–2386. https://doi.org/10.1145/2396761.2398646

Marshakova-Shaikevich, I. (1973). System of document connections based on references. *Nauchno-Tekhnicheskaya Informatsiya Seriya 2-Informatsionnye Protsessy*, *6*, 3–8.

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*, *126*(1), 871–906. https://doi.org/10.1007/s11192-020-03690-4

Milanez, D. H., Noyons, E., & de Faria, L. I. L. (2016). A delineating procedure to retrieve relevant publication data in research areas: The case of nanocellulose. *Scientometrics*, *107*(2), 627–643. https://doi.org/10.1007/s11192-016-1922-5

Morillo, F., Bordons, M., & Gómez, I. (2001). *An approach to interdisciplinarity through bibliometric indicators*. 20.

Morris, S. A., & van der Veer Martens, B. (2008). Mapping research specialties. *Annual Review of Information Science and Technology*, *42*(1), 213–295. https://doi.org/10.1002/aris.2008.1440420113

Munim, Z. H., Dushenko, M., Jimenez, V. J., Shakil, M. H., & Imset, M. (2020). Big data and artificial intelligence in the maritime industry: A bibliometric review and future research directions. *Maritime Policy & Management*, *47*(5), 577–597. https://doi.org/10.1080/03088839.2020.1788731

Muñoz-Écija, T., Vargas-Quesada, B., & Chinchilla Rodríguez, Z. (2019). Coping with methods for delineating emerging fields: Nanoscience and nanotechnology as a case study. *Journal of Informetrics*, *13*(4), 100976. https://doi.org/10.1016/j.joi.2019.100976

Narin, F., Carpenter, M., & Berlt, N. C. (1972). Interrelationships of scientific journals. *Journal of the American Society for Information Science*, *23*(5), 323–331. https://doi.org/10.1002/asi.4630230508

Newman, M. E. J. (2004a). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, *101*(suppl 1), 5200–5205. https://doi.org/10.1073/pnas.0307545100

Newman, M. E. J. (2004b). Fast algorithm for detecting community structure in networks. *Physical Review E*, *69*(6), 066133. https://doi.org/10.1103/PhysRevE.69.066133

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 026113. https://doi.org/10.1103/PhysRevE.69.026113

Oberski, J. E. J. (1988). Chapter 14—Some statistical aspects of co-citation cluster analysis and a judgment by physicists. In A. F. J. Van raan (Ed.), *Handbook of Quantitative Studies of Science and Technology* (pp. 431–462). Elsevier. https://doi.org/10.1016/B978-0-444-70537-2.50019-2

Peel, L., Larremore, D. B., & Clauset, A. (2017). The ground truth about metadata and community detection in networks. *Science Advances*, *3*(5), e1602548. https://doi.org/10.1126/sciadv.1602548

Perianes-Rodriguez, A., & Ruiz-Castillo, J. (2017). A comparison of the Web of Science and publication-level classification systems of science. *Journal of Informetrics*, *11*(1), 32–45. https://doi.org/10.1016/j.joi.2016.10.007

Persson, O. (1994). The intellectual base and research fronts of JASIS 1986–1990. *Journal of the American Society for Information Science*, *45*(1), 31–38. https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<31::AID-ASI4>3.0.CO;2-G

Porter, A., Cohen, A., Roessner, J. D., & Perreault, M. (2007). Measuring researcher interdisciplinarity. *Scientometrics*, *72*(1), 117–147. https://doi.org/10.1007/s11192-007-1700-5

Purnell, P. J. (2022). A comparison of different methods of identifying publications related to the United Nations Sustainable Development Goals: Case study of SDG 13—Climate Action. *Quantitative Science Studies*, *3*(4), 976–1002. https://doi.org/10.1162/qss_a_00215

Raan, A. F. J. van. (2005a). Measurement of Central Aspects of Scientific Research: Performance, Interdisciplinarity, Structure. *Measurement: Interdisciplinary Research and Perspectives*, *3*(1), 1–19. https://doi.org/10.1207/s15366359mea0301_1

Raan, A. F. J. van. (2005b). Measuring Science. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.), *Handbook of Quantitative Science and Technology Research* (pp. 19–50). Springer Netherlands. http://link.springer.com.focus.lib.kth.se/chapter/10.1007/1-4020-2755-9_2

Rafols, I., & Leydesdorff, L. (2009). Content-based and algorithmic classifications of journals: Perspectives on the dynamics of scientific communication and indexer effects. *Journal of the American Society for Information Science and Technology*, *60*(9), 1823–1835. https://doi.org/10.1002/asi.21086

Rafols, I., Noyons, E., Confraria, H., & Ciarli, T. (2021). *Visualising plural mappings of science for Sustainable Development Goals (SDGs)*. SocArXiv. https://doi.org/10.31235/osf.io/yfqbd

Rafols, I., Porter, A. L., & Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, *61*(9), 1871–1887. https://doi.org/10.1002/asi.21368

Reichardt, J., & Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E*, *74*(1), 016110. https://doi.org/10.1103/PhysRevE.74.016110

Robert K. Merton. (1973). *The sociology of science: Theoretical and empirical investigations*. Univof Chicago Pr.

Robertson, S. E., & Jones, K. S. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science*, *27*(3), 129–146. https://doi.org/10.1002/asi.4630270302

Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. *The European Physical Journal Special Topics*, *178*(1), 13–23. https://doi.org/10.1140/epjst/e2010-01179-1

Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(4), 1118–1123. https://doi.org/10.1073/pnas.0706851105

Ruiz-Castillo, J., & Waltman, L. (2015). Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *Journal of Informetrics*, *9*(1), 102–117. https://doi.org/10.1016/j.joi.2014.11.010

Ruiz-Real, J. L., Uribe-Toril, J., Torres Arriaza, J. A., & de Pablo Valenciano, J. (2020). A Look at the Past, Present and Future Research Trends of Artificial Intelligence in Agriculture. *Agronomy-Basel*, *10*(11), 1839. https://doi.org/10.3390/agronomy10111839

Same data—different results? Towards a comparative approach to the identification of thematic structures in science. (2017). *Scientometrics*, *111*(2), 979–979. https://doi.org/10.1007/s11192-017-2295-0

Scharnhorst, A., Börner, K., & Besselaar, P. (2012). *Models of Science Dynamics*. Springer Berlin Heidelberg.

Schubert, A., & Soós, S. (2010). Mapping of science journals based on h-similarity. *Scientometrics*, *83*(2), 589–600. https://doi.org/10.1007/s11192-010-0167-y

Schuemie, M. J., Talmon, J. L., Moorman, P. W., & Kors, J. A. (2009). Mapping the domain of medical informatics. *Methods of Information in Medicine*, *48*(1), 76–83.

Sjögårde, P. (2022a). *PubMed Classification*. https://figshare.com/collections/PubMed_Classification/5610971

Sjögårde, P. (2022b). Exploring user needs in relation to algorithmically constructed classifications of publications: A case study. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *Proceedings of the 26thInternational Conference on Science and Technology Indicators*. https://doi.org/10.5281/zenodo.6959252

Sjögårde, P. (2022c). Improving overlay maps of science: Combining overview and detail. *Quantitative Science Studies*, 1–40. https://doi.org/10.1162/qss_a_00216

Sjögårde, P., & Didegah, F. (2022). The association between topic growth and citation impact of research publications. *Scientometrics*, *127*(4), 1903–1921. https://doi.org/10.1007/s11192-022-04293-x

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*(4), 265–269. https://doi.org/10.1002/asi.4630240406

Small, H., Boyack, K. W., & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy*, *43*(8), 1450–1467. https://doi.org/10.1016/j.respol.2014.02.005

Small, H. G., & Koenig, M. E. D. (1977). Journal clustering using a bibliographic coupling method. *Information Processing & Management*, *13*(5), 277–288. https://doi.org/10.1016/0306-4573(77)90017-6

Small, H., & Griffith, B. C. (1974). The Structure of Scientific Literatures I: Identifying and Graphing Specialties. *Science Studies*, *4*(1), 17–40.

Small, H., & Sweeney, E. (1985). Clustering thescience citation index using co-citations. I. A comparison of methods. *Scientometrics*, *7*(3), 391–409. https://doi.org/10.1007/BF02017157

Small, H., Sweeney, E., & Greenlee, E. (1985). Clustering the science citation index using co-citations. II. Mapping science. *Scientometrics*, *8*(5), 321–340. https://doi.org/10.1007/BF02018057

Smiraglia, R. P., & van den Heuvel, C. (2013). Classifications and concepts: Towards an elementary theory of knowledge interaction. *Journal of Documentation; Bradford*, *69*(3), 360–383. http://dx.doi.org.ezproxy.its.uu.se/10.1108/JD-07-2012-0092

Šubelj, L., van Eck, N. J., & Waltman, L. (2016). Clustering scientific publications based on citation relations: A systematic comparison of different methods. *PLOS ONE*, *11*(4), e0154404. https://doi.org/10.1371/journal.pone.0154404

Sugimoto, C. R., & Weingart, S. (2015). The kaleidoscope of disciplinarity. *Journal of Documentation*, *71*(4), 775–794. https://doi.org/10.1108/JD-06-2014-0082

Tahamtan, I., & Bornmann, L. (2022). The Social Systems Citation Theory (SSCT): A proposal to use the social systems theory for conceptualizing publications and their citations links. *Profesional de La Información*, *31*(4), Article 4. https://doi.org/10.3145/epi.2022.jul.11

Takeda, Y., & Kajikawa, Y. (2009). Optics: A bibliometric approach to detect emerging research domains and intellectual bases. *Scientometrics*, *78*(3), 543–558. https://doi.org/10.1007/s11192-007-2012-5

Takeda, Y., & Kajikawa, Y. (2010). Tracking modularity in citation networks. *Scientometrics*, *83*(3), 783–792. https://doi.org/10.1007/s11192-010-0158-z

Takeda, Y., Mae, S., Kajikawa, Y., & Matsushima, K. (2009). Nanobiotechnology as an emerging research domain from nanotechnology: A bibliometric approach. *Scientometrics*, *80*(1), 23–38. https://doi.org/10.1007/s11192-007-1897-3

Traag, V. A., Dooren, P., & van Nesterov, Y. (2011). Narrow scope for resolution-limit-free community detection. *Physical Review E*, *84*(1), 016114. https://doi.org/10.1103/PhysRevE.84.016114

Traag, V. A., Waltman, L., & Eck, N. J. van. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, *9*(1), 5233. https://doi.org/10.1038/s41598-019-41695-z

Treeratpituk, P., & Callan, J. (2006). Automatically labeling hierarchical clusters. *Proceedings of the 2006 National Conference on Digital Government Research*, 167–176. https://doi.org/10.1145/1146598.1146650

van Eck, N. J. van, Waltman, L., Dekker, R., & Berg, J. van den. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology*, *61*(12), 2405–2416. https://doi.org/10.1002/asi.21421

van Raan, A. F. J. (1998). In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics*, *43*(1), 129–139. https://doi.org/10.1007/BF02458401

Velden, T., Boyack, K. W., Gläser, J., Koopman, R., Scharnhorst, A., & Wang, S. (2017). Comparison of topic extraction approaches and their results. *Scientometrics*, *111*(2), 1169–1221. https://doi.org/10.1007/s11192-017-2306-1

Velden, T., Yan, S., & Lagoze, C. (2017). Mapping the cognitive structure of astrophysics by infomap clustering of the citation network and topic affinity analysis. *Scientometrics*, *111*(2), 1033–1051. https://doi.org/10.1007/s11192-017-2299-9

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 1–22. https://doi.org/10.1162/qss_a_00112

Wallace, M. L., Gingras, Y., & Duhon, R. (2009). A new approach for detecting scientific specialties from raw cocitation networks. *Journal of the American Society for Information Science and Technology*, *60*(2), 240–246. https://doi.org/10.1002/asi.20987

Waltman, L., Boyack, K. W., Colavizza, G., & van Eck, N. J. (2020). A principled methodology for comparing relatedness measures for clustering publications. *Quantitative Science Studies*, *1*(2), 691–713. https://doi.org/10.1162/qss_a_00035

Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, *63*(12), 2378–2392. https://doi.org/10.1002/asi.22748

Waltman, L., & van Eck, N. J. (2013a). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, *96*(3), 699–716. https://doi.org/10.1007/s11192-012-0913-4

Waltman, L., & van Eck, N. J. (2013b). A smart local moving algorithm for large-scale modularity-based community detection. *The European Physical Journal B*, *86*(11), 471. https://doi.org/10.1140/epjb/e2013-40829-0

Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, *4*(4), 629–635. https://doi.org/10.1016/j.joi.2010.07.002

Wang, L., Topaz, M., Plasek, J. M., & Zhou, L. (2017). Content and Trends in Medical Informatics Publications over the Past Two Decades. *Studies in Health Technology and Informatics*, *245*, 968–972.

Wang, Q. (2018). A bibliometric model for identifying emerging research topics. *Journal of the Association for Information Science and Technology*, *69*(2), 290–304. https://doi.org/10.1002/asi.23930

Wang, Q., & Ahlgren, P. (2018). Measuring the interdisciplinarity of research topics. In *STI 2018 Conference Proceedings* (pp. 134–142). https://openaccess.leidenuniv.nl/handle/1887/65325

Wang, Q., & Jeppsson, T. (2022). Identifying benchmark units for research management and evaluation. *Scientometrics*, *127*(12), 7557–7574. https://doi.org/10.1007/s11192-022-04413-7

Wang, Q., & Schneider, J. W. (2019). Consistency and validity of interdisciplinarity measures. *Quantitative Science Studies*, *1*(1), 239–263. https://doi.org/10.1162/qss_a_00011

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics*, *10*(2), 347–364. https://doi.org/10.1016/j.joi.2016.02.003

Wedell, E., Park, M., Korobskiy, D., Warnow, T., & Chacko, G. (2022). Center–periphery structure in research communities. *Quantitative Science Studies*, *3*(1), 289–314. https://doi.org/10.1162/qss_a_00184

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, *32*(3), 163–171. https://doi.org/10.1002/asi.4630320302

White, H., & McCain, K. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of The American Society for Information Science and Technology*, *49*(4), 327–355.

Wouters, P. (1999). Beyond the holy grail: From citation theory to indicator theories. *Scientometrics*, *44*(3), 561–580. https://doi.org/10.1007/BF02458496

Xu, S., Hao, L., An, X., Pang, H., & Li, T. (2020). Review on emerging research topics with key-route main path analysis. *Scientometrics*, *122*(1), 607–624. https://doi.org/10.1007/s11192-019-03288-5

Yan, E., Ding, Y., & Jacob, E. K. (2012). Overlaying communities and topics: An analysis on publication networks. *Scientometrics*, *90*(2), 499–513. https://doi.org/10.1007/s11192-011-0531-6

Yan, E., Ding, Y., Milojević, S., & Sugimoto, C. R. (2012). Topics in dynamic research communities: An exploratory study for the field of information retrieval. *Journal of Informetrics*, *6*(1), 140–153. https://doi.org/10.1016/j.joi.2011.10.001

Yan, E., Ding, Y., & Zhu, Q. (2010). Mapping library and information science in China: A coauthorship network analysis. *Scientometrics*, *83*(1), 115–131. https://doi.org/10.1007/s11192-009-0027-9

Yun, J., Ahn, S., & Lee, J. Y. (2020). Return to basics: Clustering of scientific literature using structural information. *Journal of Informetrics*, *14*(4), 101099. https://doi.org/10.1016/j.joi.2020.101099

Zeng, A., Shen, Z., Zhou, J., Fan, Y., Di, Z., Wang, Y., Stanley, H. E., & Havlin, S. (2019). Increasing trend of scientists to switch between topics. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-11401-8

Zhang, L., Liu, X., Janssens, F., Liang, L., & Glänzel, W. (2010). Subject clustering analysis based on ISI category classification. *Journal of Informetrics*, *4*(2), 185–193. https://doi.org/10.1016/j.joi.2009.11.005