Faculty of Social Sciences
Department of Sociology
Centre for Research and Development monitoring (ECOOM)

Disciplines, specialization and interdisciplinarity in the social sciences and humanities

Disciplines, specialisatie en interdisciplinariteit in de sociale en humane wetenschappen

PhD thesis submitted for the degree of doctor in library and information science at the University of Antwerp to be defended by Joshua Eykens

Supervisors:

Prof. Dr. Tim C. E. Engels
Dr. Raf Guns

Antwerp, 2022

Jury members

Prof. Dr. Frédéric Vandermoere (chair of the doctoral committee and of the jury)

Prof. Dr. Ismael Rafols

Prof. Dr. Lin Zhang

Prof. Dr. Mike Kestemont

Dr. Raf Guns

Prof. Dr. Raf Vanderstraeten

Prof. Dr. Tim C. E. Engels

Table of Contents

List of figures

List of tables

Acknowledgements

I would like to explicitly thank my two supervisors, Raf Guns and Tim Engels. You have both given me room to think and wander over the course of my PhD trajectory. Raf I would like to thank you in particular, for your expert guidance on conceptual and methodological questions, but also for your technical assistance with programming problems. Without your proficiency in informetric concepts and swift programming skills, this thesis would have looked a lot different. Tim, I'm grateful for the vivid discussions you initiated around the subjects close to my PhD research, but also on topics at a larger distance. For stimulating me and giving me the opportunity and the time necessary to also study other things than what is presented in this thesis.

I would also like to express gratitude to Raf Vanderstraeten, Linda Sīle, Frédéric Vandermoere, Ronald Rousseau, and Pieter Spooren. Pieter, it was a pleasure being your neighbor in the E-building. Thank you for the motivating words from time to time. Ronald, you have been an inspiration along the way of my PhD career. Seeing a researcher doing what he does best with that much passion is truly inspiring. Frédéric, I want to thank you for the interesting discussions we have had during formal and informal meetings. Linda, you have had a very large influence on my professional identity as a researcher. I consider you an example of how a social scientist should think, work and write. Raf, you have been there right from the start of my training as a sociologist. It is a pleasure to still be able to exchange ideas and texts with you, and I would like to thank you very much for that. Eline, Cristina, Hongyu and Peter, thanks for being wonderful people. Also, many thanks to the staff of the Antwerp Doctoral School, for being very pleasant and helpful office colleagues.

I am also grateful for the international experience I have gained over the past four years. Thank you to all the wonderful colleagues I have met abroad, but especially Janne Pölönen, Hanna-Mari Puuska and Vidar Røeggen.

Thank you to my dear friends and family, who from time to time reminded me that there is more to life than work: Bart, Bob, Dries, Florian, Frank, Freek, Geert, Jan, Jonathan, Maarten, Mathias, Mathijs, Piet, Pieter, Raf, Ruben, Seppe, Tom, Wesley, and Wout, for the meals, drinks, and fantastic moments we shared. I would also like to explicitly thank my mother, for her love and all the support she has offered throughout my (rather long) career as a university student.

Carolien, I could never have completed all this work without you. I couldn't have imagined a better partner. During the past few months, however, someone else has joined our household as well.

I would like to dedicate this thesis to you, my two great sources of love, Carolien and Fons.

Abstract (Dutch)

Dit proefschrift heeft als doel om specialisering en interdisciplinariteit in de sociale en humane wetenschappen in kaart te brengen. Interdisciplinariteit neemt al geruime tijd een centrale plaats in in het beleidsdiscours en zal binnenkort ook in de vorm een parameter toegevoegd worden in het allocatiemodel voor Vlaamse onderzoeksfinanciering. Het centrale idee dat naar voor wordt geschoven om deze keuze te kracht bij te zetten, is dat de complexe problemen waar we als samenleving mee te kampen hebben enkel opgelost kunnen worden wanneer kennis afkomstig uit verschillende vakgebieden wordt gecombineerd. Ecologische en maatschappelijke crises, zoals de door mensen versnelde klimaatopwarming, politiek extremisme, of de veelzijdige problematiek van een wereldwijde pandemie vragen een geïntegreerde aanpak, dewelke maar moeilijk door een enkele discipline geboden kan worden, zo luidt het.

In de eerste twee hoofdstukken van deze scriptie gaan we eerst dieper in op wetenschappelijke specialisering en de veranderende positie van onderzoeksdisciplines in de wetenschappen in het algemeen. We schuiven een visie naar voor waarin centraal staat dat door een toegenomen interne en externe differentiatie de disciplinaire structuur van het wetenschapssysteem onderhevig is aan verandering. Differentiatie leidt bijvoorbeeld tot het ontstaan van diverse specialismen dewelke op hun beurt voortdurend de grenzen van disciplines aftasten. Binnen traditionele disciplines, zoals bijvoorbeeld de psychologie of de economie, kunnen we als gevolg van de enorme groei van het wetenschapssysteem het ontstaan van talloze subdisciplines en kleinschaligere specialismen optekenen. Deze ontstaan veelal rond nieuwe onderwerpen, theorieën of methodes waarrond of waarmee onderzoek uitgevoerd wordt. Maar ook door impulsen extern aan het wetenschapssysteem ontstaan nieuwe specialismen. Het 'verwetenschappelijken' van sociale nieuwigheden of opkomend onderzoek naar de zich snel ontwikkelende maatschappelijke deelsystemen leidt tot het ontstaan van domeinen die geen directe of net meerdere disciplinaire voorganger(s) kennen (bv. new media studies, scientometrics, etc.). Deze specialismen functioneren als de primaire referentiegroep voor onderzoekers, en zoals talloze gevalstudies aantonen zijn ze per

definitie interdisciplinair. We besteden ook aandacht aan transdisciplinariteit. Hoe leidt toegenomen externe differentiatie tot samenwerkingen met actoren buiten het academische systeem?

De voornaamste les die uit deze inleidende hoofdstukken voortvloeit heeft zowel betrekking op het onderzoek naar interdisciplinariteit als onderzoeksevaluatie. We halen aan dat een herwaardering van de categorieën 'discipline' en 'specialisering' noodzakelijk zijn binnen bibliometrisch onderzoek. De geschiedenis leert ons dat de toegenomen specialisering en differentiatie de tegenstelling disciplinair-interdisciplinair irrelevant maakt. In navolging daarvan worden richtlijnen voor onderzoeksevaluatie die betrekking heeft op interdisciplinariteit samengevat.

Het tweede luik van deze scriptie behandelt de bibliometrische operationalisering van specialismen in de SHW. Gangbare benaderingswijzen zijn veelal gebaseerd op citatie-analyses, maar voor de SHW is dit een in verschillende opzichten problematische benadering te noemen. Op zoek naar een alternatief reconstrueerden we daarom aan de hand van machine-learningtechnieken en tekstuele gegevens uit disciplinespecifieke databanken (Sociological Abstracts, EconLit, en ERIC – Education Resources Information Center) een fijnmazige onderzoeksclassificatie op het niveau van subdisciplines en specialismen. Tezelfdertijd lieten we experten een reeks wetenschappelijke publicaties uit dezelfde set classificeren en bestudeerden we vervolgens de mate van consistentie van de toegewezen categorieën. Hieruit kwam naar voor dat de implementatie van een classificatie op het niveau van onderzoeksspecialismen een moeilijke opdracht is, zowel in het geval van geautomatiseerde benaderingen op basis van machine learning als voor domein experten uit de betrokken disciplines. In een derde studie kijken we naar de uitkomsten van clusteranalyse om documenten te groeperen op basis van tekstuele gelijkenissen. De conclusie is dat recent geïntroduceerde vectorisatietechnieken op basis van neurale netwerken beloftevolle uitkomsten kunnen bieden.

Aan de hand van een omvattende bibliografische databank voor de Vlaamse SHW, VABB-SHW, bestuderen we in het derde en laatste deel disciplinariteit en interdisciplinariteit. De twee empirische studies die gepresenteerd worden behandelen opeenvolgend de interdisciplinaire mobiliteit van onderzoekers en identiteit van onderzoeksspecialismen respectievelijk. In overeenkomst met hetgeen we suggereren in de inleidende hoofstukken kunnen we voor zowel de onderzoekers als de specialismen maar in zeer beperkte mate (of helemaal geen in het geval van de specialismen) spreken van strikt disciplinaire identiteiten of profielen. Onderzoekers in de SHW publiceren bijvoorbeeld in toenemende mate in uiteenlopende disciplines. We tonen tevens aan dat diezelfde onderzoekers niet enkel publiceren overheen discipline categorieën die inhoudelijk sterk op elkaar lijken. We merken bovendien op dat voor degenen die over een groot aantal verschillende categorieën publiceren er maar in beperkte mate sprake is van cognitieve mobiliteit. Het lijkt erop dat deze disciplinair mobiele – of interdisciplinaire groep van onderzoekers in feite hyper-gespecialiseerd werk verricht. Voor de onderzoeksspecialismen in de SHW op hun beurt, stellen we vast dat geen enkel specialisme strikt gedisciplineerd is. Op basis van een contrastering van de disciplinaire diversiteit op vlak van de affiliatie van auteurs en de classificatie van de onderzoeksoutput stellen we een typologie op van verschillende vormen van interdisciplinaire specialismen.

Concrete aanbevelingen voor bibliometrisch onderzoek naar interdisciplinariteit, meer specifiek in de context van de SHW worden geformuleerd, alsook een reeks bedenkingen die relevant kunnen zijn voor degenen die bevoegd zijn met onderzoeksevaluatie of beleid.

Abstract (English)

Is the unity of the sciences threatened to be lost by increased specialization? Do we think too much in boxes? Is there a need for more interdisciplinarity? Although interdisciplinarity has occupied a central place in policy discourse for some time now, a parameter for it will soon also be added in the allocation model for Flemish research funding. The central idea put forward to substantiate this choice is that the complex problems we face as a society can only be solved when knowledge from different disciplines is combined. Ecological and social crises, such as human-accelerated climate warming, political extremism, or the multifaceted problems of a global pandemic, require an integrated approach, which is difficult to offer by a single discipline, it is said.

In the first two chapters of this thesis, we first examine scientific specialization and the changing position of research disciplines in science in general. Due to an increased internal and external differentiation, the disciplinary structure of the scientific system is subject to change. For example, internal differentiation leads to the emergence of various specialisms, which in turn continuously explore the boundaries of disciplines. Within traditional disciplines, such as psychology or economics, we can note the emergence of countless subdisciplines and smaller-scale specialisms as a result of the enormous growth of the scientific system. These often arise around new subjects, theories or methods around which or with which research is carried out. But new specialisms are also created as a result of impulses external to the science system. The 'scientificization' of social novelties or emerging research into the rapidly developing social subsystems leads to the emergence of domains that have no direct or just several disciplinary predecessor(s) (e.g. new media studies, scientometrics, etc.). These specialties function as the primary reference group for researchers, and as numerous case studies demonstrate, they are interdisciplinary by definition. We also pay attention to transdisciplinarity. How does increased external differentiation lead to collaborations with actors outside the academic system?

The main lesson that emerges from these introductory chapters relates to research on interdisciplinarity as well as research evaluation. We point out that a revaluation of the categories 'discipline' and 'specialization' is necessary within bibliometric research. History teaches us that increased specialization and differentiation makes the disciplinary-interdisciplinary opposition irrelevant. Following this, guidelines for research evaluation related to interdisciplinarity are summarized.

The second part of this thesis deals with the bibliometric operationalization of specialties in the social sciences and humanities (SSH). Common methods are often based on citation analyses, but for the SSH this is a problematic approach in several respects. In search of an alternative, we therefore reconstructed a fine-grained research classification at the level of subdisciplines and specialisms using machine learning techniques and textual data from discipline-specific databases (Sociological Abstracts, EconLit, and ERIC – Education Resources Information Center). At the same time, we had experts classify a series of scientific publications from the same set and then studied the degree of consistency of the assigned categories. This showed that the implementation of a classification at the level of research specialties is a difficult task, both in the case of algorithmic approaches based on machine learning and for domain experts from the involved disciplines. In a third study, we look at the results of cluster analysis to group documents based on textual similarities. The conclusion is that recently introduced vectorization techniques based on neural networks can offer promising results.

On the basis of a comprehensive bibliographic database for the Flemish SSH (VABB-SHW), we study disciplinarity and interdisciplinarity in the third and final part. The two empirical studies presented address the interdisciplinary mobility of researchers and identity of research specialisms, respectively. In accordance with what we suggest in the introductory chapters, we can only speak to a very limited extent (or none at all in the case of the specialisms) of strictly disciplinary identities or profiles for both the researchers and the specialties. Researchers in the SSH, for example, are increasingly publishing in a variety of disciplines. We also show that the same researchers do not only publish across discipline categories that are very similar in content. We also note that for those who publish across a large number of different categories, there is only a limited degree of cognitive mobility. It seems that this disciplinary mobile - or interdisciplinary group of researchers is in fact conducting hyper-specialized work. For the research specialisms in the SSH in turn, we note that no specialism is strictly disciplined. Based on contrasting the disciplinary diversity in terms of author affiliation and the classification of research output, we establish a typology of different forms of interdisciplinary specialisms.

Concrete recommendations for bibliometric research into interdisciplinarity, more specifically in the context of the SSH are formulated, as well as a series of considerations that may be relevant for those in charge of research evaluation or policy.

1 Introduction

Interdisciplinary research is commonly associated with or used as a synonym for scientific and technological innovation (Ledford, 2015). Revolutionary discoveries and developments made at the boundaries of different fields are indeed omnipresent in our recent history of science. For instance, progress in nanotechnology led to the development of computing devices which we now cherish as if they have always been with us, and advances in physics led to the development of magnetic resonance imaging. The latter has brought about many breakthroughs in medical diagnostics and also in neuroscience (Viard, Eustache, & Segobin, 2021). More recently, revolutionary molecular and chemical biology research into the possibilities for synthesizing mRNA (starting in the 70s and 80s) has resulted in the development of COVID19 vaccines (Dolgin, 2021).

It is beyond doubt that none of these technological and scientific milestones could have been achieved in an academic environment in which disciplinary communities of scientists exist as static and entirely secluded entities. Dynamic interaction between researchers from across and within different specializations was a prerequisite for each, not only in the actual development of, but also in the run up to those events.

Contemporary discourse on interdisciplinary research (or IDR in brief), however, gives us the impression that it is the only possible way forward for complex problem solving, that it is a necessary prerequisite. Science is even perceived to be entering a new era, in which the disciplines will soon become irrelevant. Within this perspective, purely disciplinary research is regarded as being antecedent to IDR, impairing innovative scientific research. In fact, new modes of science have been proposed that could well be replacing the traditional disciplinary system with an integrated counterpart (for recent suggestions, see post-disciplinarity, mode-2 science, etc.) (Graff, 2016; Weingart, 2016).

For the societal issues we face today, IDR has indeed been and will be useful. All of the sustainable development goals (SDGs) require insights from different disciplinary angles to be tackled adequately

(Keynejad, Yapa, & Ganguli, 2021). The climate crisis, for example, is a global-scale emergency which cannot be addressed from, say, a purely economic perspective. More scientific impact and relevant technological breakthroughs on pressing issues are expected if we are able to transcend the boundaries of traditional disciplines and further integrate the social sciences and humanities.

1.1 A definition of interdisciplinarity

Before we delve deeper into its perceived relevance and importance, let us first turn to a simple definition of interdisciplinarity proposed by the National Academy of Sciences (2004), which is widely cited in science policy and research into IDR (also see chapter 3);

> *"Interdisciplinary research is a mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice"*

As stated above and also apparent from this definition is the aspect of (two or more) disciplines. A coordinated integration of the knowledge resources developed by different disciplines is considered a central function of interdisciplinary research. The second part of this definition highlights the promise of IDR for complex problem solving. The implicit assumption that this can only take place if the boundaries of disciplines are transcended is repeated here as well. As we will see further on, however, the straightforward appearance of this definition should not be mistaken. Many conceptual and operational complexities exist, also on the level of disciplines. How do we define and delineate a discipline (Hammarfelt, 2019)? Where do the disciplinary boundaries lie? And what is interdisciplinary knowledge integration (Clark & Wallace, 2015; Leydesdorff & Ivanova, 2020)?

1.2 Relevance of the topic

According to many, interdisciplinarity is a mode of research which runs against the tide of traditional disciplinary research and therefore requires extra efforts (and funds). And indeed, given these large promises IDR is assumed to bring about, it is lauded by science policy makers across the globe (Rylance, 2015). From the National Natural Science Foundation of China to the National Science Foundation of the US and the European Research Council (Allmendinger, 2015), all give high priority to IDR in their funding calls and mission statements. In Flanders (the Dutch speaking part of Belgium) too this line of thought about the importance of IDR has reached policy circles. The Flemish government decided upon the instalment of a new funding parameter taking into account IDR for the distribution of government block funding over the five Flemish universities from 2024 onwards (see BOF besluit 2018[1]) (Luwel, 2021).

Individual universities themselves also have a history of organizing and encouraging interdisciplinarity (Jacobs & Frickel, 2009). At Ghent University (Flanders, Belgium), for example, we can now find 10 specific research consortia[2] (IOCs or 'Interdisciplinaire Onderzoeksconsortia' in Dutch) which were introduced to develop interdisciplinary science on topics such as globalisation, crime and mental health (Klima, Meysman, Carlier, Dewaele, & de Smet, 2019). For the funding of these units, a devoted financing mechanism has been initiated. Around the same time (2017-2018), the oldest university in Flanders, KU Leuven, launched a policy plan to further spur IDR[3]. At the University of Antwerp, CeMIS is an example of an interdisciplinary research centre, focusing on migration and intercultural studies. Additionally, each of the 15 centres of excellence at UAntwerp includes professors of at least 3 research groups to facilitate interdisciplinary interaction. The other universities also make great efforts and calls to enhance

---

[1] See https://www.ewi-vlaanderen.be/nieuws/vlaamse-regering-keurt-bof-besluit-definitief-goed-35-miljoen-euro-extra-voor-universiteiten
[2] https://www.ugent.be/nl/onderzoek/maatschappij/idc/overzicht.htm
[3] https://nieuws.kuleuven.be/en/content/2018/policy-plan-interdisciplinarity

interdisciplinary research, be it by establishing new interdisciplinary research centres, interdisciplinary curricula, or funding opportunities.

The integration of social sciences and humanities (SSH) and science, technology, engineering, and mathematics (STEM) disciplines is a specific theme which is encouraged in science policy. SSH disciplines are often absent in (IDR) funding calls (Pedersen, 2016) but are nevertheless necessary for solving the most complex societal problems (Viseu, 2015). The European Commission in this light stated as one of its missions for Horizon 2020 the integration of social sciences and humanities[4] in each of the priorities (Graf, 2019). But where to start? How can we locate disciplines and the development of interdisciplinarity?

1.3 Science studies and interdisciplinarity

Given all the attention raised and money invested by science policy makers and universities alike, it should not surprise that IDR is currently attracting quite some attention within the field of science studies. This is also observable in the form of a rapid surge of the relevant scholarly literature on the topic from 2006 onwards (see Larivière & Gingras, 2014, pp. 188-190). Research into the development of disciplines and interdisciplinarity is carried out in several fields. Philosophers, sociologists, and historians of science for example have since long devoted a great amount of work on theoretical and historical aspects of disciplines (see for example Stichweh, 1992). Today a central field working on issues relating to indicating and measuring IDR is bibliometrics or, more specifically, scientometrics. In this field, quantitative indicators are designed and used to measure and characterize disciplines, teams, researchers, or documents and their degree of interdisciplinarity.

By applying scientometric methods and mathematical models, communication processes in science can be studied. Citation and collaboration patterns between different disciplines, knowledge importing, exporting, and integration by and from different fields (see chapter 3), or the structural evolution of research specialties or

[4] See https://ec.europa.eu/info/research-and-innovation/research-area/social-sciences-and-humanities/ssh-integration_en

disciplines (see chapter 8) are just a few of the broad topics studied by scientometricians. In science policy settings, the indicators developed within this field are the primary tools used to assess and locate interdisciplinarity.

Within science studies, IDR indicators have been deployed to investigate whether this type of research has more citation impact (Larivière & Gingras, 2010), what the different aspects of IDR are which lead to higher citation impact (Wang, Thijs, & Glänzel, 2015; Yegros-Yegros, Rafols, & D'Este, 2015), how IDR is evolving across different disciplinary communities in the SSH (Zhou, Guns, & Engels, 2021), if it leads to novelty (Fontana, Iori, Montobbio, & Sinatra, 2020), etc. Although regarded as beneficial in many aspects, on the one hand there exists some evidence that interdisciplinary project proposals have decreased chances in obtaining funding (Bromham, Dinnage, & Hua, 2016) mainly because peer review practices in these contexts are discipline-specific. Seeber, Vlegels, and Cattaneo (2022) on the other have recently refuted this claim.

On a more fundamental level, the bibliometric community shares a general interest with sociologists and historians of sciences for how and where disciplinary and interdisciplinarity research develops. The dominant methodological frameworks that are used in these studies rely on citation measurement and diversity calculations based on predefined disciplinary classifications of journals and researchers (Wagner et al., 2011). While at first sight this might seem as a reasonable approach, this line of research has been shown to be problematic in different ways.

1.4 State of the art

Bibliometric studies of interdisciplinarity range from case-studies detailing the inter-disciplinary nature of individual fields or disciplines to the development of indicators deemed fit for measuring the degree of interdisciplinarity, on the level of individual researchers, papers, journals, fields and science as a whole[5].

---

[5] In this section I will only briefly summarize the literature on indicator development. A more extensive discussion can be found in chapter 3.

1.4.1 *Indicating interdisciplinarity*

Within the bibliometric community, the above-mentioned definition of IDR (p. 15, this chapter) is a standard reference (Glänzel & Debackere, 2021, p. 2). With the integrative component of interdisciplinary research being the essence. Glänzel and Debackere (2021) further highlight the two main paths taken in indicator development: (i) the cognitive and the (ii) organizational approach. The first centers around information flow across different units of analyses, and the second aims at capturing collaboration patterns.

The cognitive approach is the most developed of the two and is based on the assumption that the integration of knowledge can be measured by quantifying information flow into a new, interdisciplinary environment. A paper which cites references originating from different disciplinary categories, for example, is considered an interdisciplinary environment or artifact in which new relations across the disciplines are established. The paper (or any other unit carrying scientific information), it is expected, *integrates* knowledge from different disciplines. The degree of interdisciplinary then, is measured by calculating diversity indices (often Rao-Stirling diversity), ideally capturing three distinct but related properties: the number of disciplines or topics involved (variety), the balance of the distribution of disciplines or categories, and their disparity (i.e. the degree to which the categories are similar to each other in and off themselves).

For the social sciences in particular, Zhou et al. (2021) have recently studied the evolution of interdisciplinarity. They conducted large-scale diversity analyses based on citation data (from Microsoft Academic Graph) for 5 general social science disciplines over a 50-year time period. The authors found that IDR is indeed on the rise. In accordance with the findings presented by Levitt, Thelwall, and Oppenheim (2011), they further stipulate that all disciplines are becoming more interdisciplinary, but some substantial differences can be found with regard to the different aspects of diversity and their timeliness for the individual disciplines.

Interested readers are also encouraged to consult the excellent review by Rousseau, Zhang and Hu (2019).

Collaboration between researchers originating from different disciplines forms the substance of the organizational approach. Co-authorship patterns can reveal the intensity of collaborative efforts across disciplines or other units of analyses. Due to a lack of available, standardized affiliation data, this approach has not yet reached the stage of systematic inquiry. For the case of Italy, however, where detailed and structured data on this level are available, Abramo and colleagues have done considerable work in this line, also comparing the organizational approach with the cognitive one (Abramo, D'Angelo, & Di Costa, 2012; Abramo, D'Angelo, & Zhang, 2018). Abramo et al. (2018) have initiated work on automatically generating disciplinary affiliation data based on semantic information from author addresses derived from PLoS ONE. They have also studied the relationship between disciplinary diversity in reference lists (cognitive approach) and the diversity of the automatically generated affiliation data of authors.

On the level of the system of which the unit is a part, coherence is another aspect which is being measured. Coherence is a network property and can be modelled from the bottom up (Egghe & Rousseau, 2003; Rafols, 2014). It takes into account the extent to which the units in a network are related, or, in other words, form 'a meaningful constellation' (Stirling, personal communication as cited in Rafols & Meyer, 2009). If a network of papers, for example, is found to be highly clustered or 'coherent', it can be regarded as specialized. Rafols and Meyer (2009) effectively combine diversity and coherence to indicate different aspects of interdisciplinarity and specialization. A clustered set of documents in which a diversity of discipline categories is present, for example, is interpreted as an interdisciplinary specialty (p. 270).

1.4.2 Problems with current indicators of interdisciplinarity

Although indicators for IDR have reached a considerable degree of sophistication, achieving consistency between them is a difficult task. As has been shown by Q. Wang and Schneider (2019) recently, differing results can be found when one compares indicators which are supposedly measuring similar features or dimensions of the concept of IDR. The multidimensional nature of interdisciplinarity and disciplines, and the complexity underlying the process of integrating knowledge are cited as potential explanations for their contradictory findings.

Vugteveen, Lenders, and Van den Besselaar (2014) further problematize current bibliometric approaches to interdisciplinarity. The key issues they raise in reviewing methods like diversity and coherence measurement relate to (a) imposing artificial boundaries based on top-down classification, (b) the possible integration of IDR into a new field, or the emergence of new fields due to IDR are not taken into account, i.e. disciplinary change and evolution (related to a), and (c) a perceived confusion between the topological and relational perspective in network intermediation[6]. The argument which is put forth by these authors can be summarized by their following quote: 'interdisciplinarity is a temporary stage of disciplinary reconfiguration, as the further a new interdisciplinary field develops, the more disciplined it becomes. […] These changes can be observed if one avoids working with predefined fields. Definitions of research fronts and of fields and disciplines have to be dynamically based on similarities between journals and between papers.' (ibid, p. 77).

'Without referring clearly to the processes of knowledge production, communication and stabilization', van den Besselaar continues, '[…] the choice for indicators seems rather arbitrary' (van den Besselaar, 2019, p. 1). The author argues that if we want to know *how* knowledge is being integrated, we should focus our attention on processes of change in journal citation networks. Interdisciplinarity is indeed a field level dynamic, and consequently, the disciplinary or interdisciplinary identity of a field might change over time. New fields might emerge, grow, and differentiate. They could merge with other existing fields over time, decline or disappear altogether. Established and general disciplinary classification to a lesser extent take these variations into account. A relevant question in this regard

[6] To not distract from the general overview, we refer the reader to the technical discussion in the original paper by Vugteveen et al. (2014). To summarize, Vugteveen and colleagues argue that betweenness centrality should not be considered an adequate measure for knowledge brokering in information networks, or at least not in the way that it is applied by Goldstone and Leydesdorff (2006). See Vugteveen, Lenders, and van den Besselaar, 2014, p. 76, footnote 1.

becomes to what extent scientific fields, disciplines, and specialties are disciplined in and off themselves. It could for example be studied if variations in the degree of inter-disciplinarity of different fields exist at a certain point in time.

1.4.3 Mapping scientific disciplines and specialties in the SSH from the bottom up?

While the latter approaches are mostly centered around citation relations (except for the organizational approach), for the SSH this is problematic in various ways - and to a varying degree for different fields. The main concerns which can be raised for these fields relate to (i) a greater diversity in research outputs, (ii) differences in citation cultures, (iii) larger share of references to grey literature, and (iv) more locally oriented research and greater heterogeneity between fields. Bypassing these limitations can be achieved by establishing other types of relationships between the units or documents under consideration. For the social sciences and humanities, content similarity based on the actual text of publications might for example be a more suitable approach. A different logic applies here. While relations between documents are generally considered to represent a communicative process, those that are based on text or 'communicative content' can be termed the discursive approach (Morris & Van der Veer Martens, 2009, p. 230).

Following a long tradition of co-word analysis of 'problematic networks', as first initiated by Callon, Courtial, Turner, and Bauin (1983), the discursive approach has developed into a field of its own. In attempts to model scientific fields based on the co-occurrence of words or concepts, this stream of research today mostly deploys machine learning algorithms in order to reproduce existing classification schemes (what is called supervised learning) or clusters documents from the bottom up based on similarity relations between them (unsupervised learning). The application and performance of these methods has, however, only received scant attention in the case of the SSH.

1.4.4 Gaps in the literature on disciplines and interdisciplinarity involving the SSH

First it should be noted that a good deal of ambiguity is present on the conceptual level. Disciplines and IDR have been proven hard to define, they are multidimensional concepts carrying different meanings in different contexts (Sugimoto and Weingart, 2015; Hammarfelt, 2019; Huutoniemi, Klein, Bruun, & Hukkinen, 2010). For disciplines, a broad consensus exists on their multifacetedness; they have a dual identity which manifests itself in cognitive and organizational structures mirrored in the scientific system.

Journals, books and other types of research outcomes are, for example, indexed according to disciplinary classification systems, and scholars are often affiliated to departmental units which reflect disciplinary divisions. Yet, these classification systems and departmental structures are known to be context-dependent (Sīle, Guns, Vandermoere, Sivertsen, & Engels, 2021) and do not always align well (Guns, Sīle, Eykens, Verleysen, & Engels, 2018).

Universities across and within different countries have different departmental structures and classification systems developed for research output are a product of historical changes in science and society. The disciplinary structure itself is often taken for granted as a natural fact by many but is only as old as the modern system of university education. Starting in the 19th century, different disciplines have emerged and disappeared. As the scientific system grows more diverse, with numerous research specialties and synergies between disciplines emerging rapidly, the application of disciplinary classification systems deserves second thought.

The usage of universal and pre-specified classification systems could misrepresent the complexities of contemporary science may at the same time disregard the dynamics of the system. Applying measures which are solely based on such classification systems only tell us part of the story. Inconsistency on the conceptual level, i.e., how we define and delineate disciplines, or in the ways in which we approach these disciplinary systems (i.e. should disciplines be considered as fluid and contextual or static and universal entities), is leading to inconsistencies on the operational level. The standard application of ex ante disciplinary classification (like for example the Subject Categories in the Web of Science) does not take into account contextual features of disciplines or their dynamics[7]. Should they still be applied in science studies? And, if yes, how can we still make use of these systems to improve our understanding of complexities in science?

Other ways of approaching the cognitive and social structure of science have also been proposed in bibliometrics. Methodological frameworks which do not make use of a flat disciplinary ontology, without or in combination with existing classification systems. Examples are document clustering based on citation relations, detecting communities in author collaboration networks, or grouping documents based on textual similarity patterns. Methods combining these approaches have been introduced as well.

For the social sciences and humanities in particular, it is well known that researchers active in these fields often publish in multiple formats (i.e. monographs and edited volumes in addition to journal articles and conference proceedings) and languages other than English. These publication types are generally less well covered by commercial citation indexes and as such only yield partial pictures of the involvement of SSH researchers in IDR. A general question of interest here is how we can make use of bottom-up approaches like document clustering based on text to group documents from comprehensive, multi-lingual and multi-type datasets containing SSH publications.

For interdisciplinarity or IDR, the above-mentioned definition is almost quoted by default in scientometric studies. The integrative component is thereby stressed. Patterns of knowledge integration within documents are measured by calculating diversity scores based on the disciplinary categories present in reference lists. While we have already pointed out that ex-ante classification systems of disciplines do not always allow for a dynamic or fine-grained approach of the continuously specializing scientific system, these are nonetheless regarded as established natural kinds and taken for granted in many studies of IDR. For the SSH additional objections

[7] Some subject categories are added over time, but the system evolves rather slowly.

can be raised to this way of working as these fields are indeed less well covered in citation indexes.

1.5 Topic and scope of this thesis

This thesis contains scientometric studies on the boundaries of academic disciplines and interdisciplinarity involving the SSH. Research evaluation practices regarding IDR in the SSH are discussed. Methodological approaches to classify and cluster scientific documents from the SSH are explored and novel empirical observations regarding the inter- or cross-disciplinarity of research specialties and researchers from the SSH are presented. On the level of researchers, disciplinary boundary crossing or 'field mobility' is the point of interest.

The questions which I explore in this thesis can be subdivided into three overarching topics or 'parts'. The first being a conceptual one, disentangling (1) the meaning of disciplines and interdisciplinarity involving the social sciences and humanities (chapters 2 and 3). The second starts from a methodological curiosity for text-based approaches to science classification and document clustering (chapters 4, 5 and 6). For the third part I present an empirical exploration into the interplay between and the relative isolation of disciplines in the social sciences and humanities based on textual methods (chapters 7 and 8).

In a first part of this thesis, I aim to sketch a brief history of disciplines and their functions to initiate a discussion on interdisciplinary research. Different bibliometric approaches are discussed and scrutinized in terms of their applicability in a context of SSH research evaluation.

Part 1 – Disciplines and interdisciplinarity

- What are academic disciplines and how did they emerge?
    - What is their function in the contemporary academic system?
    - How are they approached in bibliometric research?
- What is interdisciplinarity, how is it understood in policy circles, and how is it assessed?

- Which indicators and procedures exist for approaching and assessing interdisciplinarity?
- Should alternatives be considered for evaluating interdisciplinarity?

The idea that overarching disciplines are still present in contemporary academia cannot be disputed. As we will discuss in chapter 2, however, in many cases they serve different functions nowadays. Grouping together communities of scholars mainly centred around subject areas or specialized topics, they provide a common reference point or address, an overarching structure. To identify the cognitive content produced by these more granular structures or fine-grained communities, I explored to what extent I could make use of text-based approaches to identify specialties in the SSH. The following research questions were asked in this regard.

Part 2 – Text-based approaches to science classification and clustering

- To what extent can we make use of supervised machine learning to reconstruct fine-grained, specialty level classifications of social sciences publications?
- What are efficient ways of text clustering to construct bottom-up classifications of social sciences and humanities publications?

For the third part discussed in this thesis I guide the reader through two empirical explorations in which I make use of text-based approaches to, on the one hand, specialties in SSH (chapter 8), and disciplines (chapter 7) on the other. I do this in order to gain insight into the interplay between disciplinary classification systems, and the roles played by specialties or research topics in this regard.

Part 3 – The interplay between and isolation of discipline categories in the social sciences and humanities

- Do researchers in the SSH publish across disciplines?
  - Do authors who switch between disciplines throughout their careers change their research direction?
  - And how does this relate to the distance they have travelled cognitively speaking?

- Do subject specialties play a role in bridging the different disciplines?
    - Which subjects are shared between disciplines and which disciplines are more open to sharing subjects with other disciplines?
    - Can we find different types of specialties in terms of their disciplinary identities?

In chapter 7 I investigate the cognitive and disciplinary mobility of authors in the social sciences and humanities. In chapter 8 I explore whether specialties operate as interdisciplinary trading grounds.

1.6 Datasets

For the methodological and empirical studies presented in this thesis I rely on data sources specifically designed for better coverage and indexing of SSH research. In part II, where I investigate document classification, data from ERIC (Education Resources Information Center), Sociological abstracts[8] and EconLit[9] are combined into a manually constructed and labeled dataset of roughly 110,000 records. ERIC[10] is the online library of education research and information and is sponsored by the US department of Education. Sociological Abstracts indexes the international literature of sociology and related behavioral sciences. EconLit is a professionally classified indexing service for economic literature. What these three databases have in common is that they are explicitly international, they cover multiple types of output and apply a professional classification based on fine-grained indexing terms derived from field specific thesauri.

Other parts (chapters 6, 7 and 8) are based on data from VABB-SHW, the Flemish Academic Bibliographic database for the Social Sciences and Humanities. VABB-SHW is a database specifically designed for better coverage of SSH research authored by scholars affiliated to a SSH research unit at one of the five Flemish Universities. Coverage starts from 2000 onwards, and includes a variety of publication types and languages (Verleysen, Ghesquière, &

[8] https://about.proquest.com/en/products-services/socioabs-set-c/
[9] https://www.aeaweb.org/econlit/
[10] https://eric.ed.gov/

Engels, 2014). Details about which subsets of the datasets are used can be found in the respective chapters.

1.7 Thesis outline

The chapters of this thesis are organized along the topics or parts presented above. The second chapter deals with the conceptual underpinnings of this thesis. Disciplines, their emergence and structure are a well-studied subject both in bibliometrics and the (historical) sociology of science. Yet some large discrepancies exist between these two strands of scholarship. I will briefly synthesize this literature and point out the commonalities and differences. I will present what I belief the main reasons for these discrepancies are, and how it imposes difficulties for bibliometric studies of science, disciplines and interdisciplinarity. Additionally, I draw attention to the importance of research specialties for this matter. These smaller yet strongly research oriented units nowadays mirror some important functions of scientific disciplines which in many bibliometric studies are still attributed to the latter.

In chapter 3 I delve deeper into the ways in which interdisciplinarity is perceived and studied, both in science policy circles and bibliometrics. I discuss the processual view on disciplines and interdisciplinarity (van den Besselaar, 2019) to further clarify where additional challenges occur in the development of bibliometric indicators for approaching interdisciplinarity (see for example Wang and Schneider, 2019). I argue that our focus should be shifted to dynamics of change in the scientific landscape in order to develop a more accurate understanding of how and where interdisciplinarity emerges. Current research assessment practices for interdisciplinarity are discussed and the seven main assessment principles introduced by Julie Thompson Klein (2008) are reiterated, with special attention for interdisciplinarity in the social sciences and humanities.

As research specialties play an increasingly important role in the scientific ecosystem, a first important task was to find and develop appropriate ways to approach these communities (part 2). Especially for the social sciences and humanities, where publication practices differ (Engels, Ossenblok, & Spruyt, 2012; Hicks, 2005) and citation information is often lacking or insufficiently covered by major bibliographic databases (Larivière, Archambault, Gingras, & Vignola-

Gagné, 2006) this becomes paramount. In chapter 4, 5 and 6 I present the results of my experiments with two text-based approaches developed for classifying and clustering documents from the social sciences and humanities on the level of research or subject specialties. The first two studies assess the suitability of supervised machine learning (or supervised document classification) to reproduce existing fine-grained classification systems. The third study (chapter 6) presents the results of an unsupervised clustering experiment in which I compare different document representation techniques.

In part 3, Chapter 7 I study the boundaries or 'boundarylessness' of disciplines by investigating the cognitive mobility of researchers in the social sciences and humanities. I show how researchers publish across disciplines throughout their careers and investigate how this relates to the cognitive distance they bridge. Chapter 8, part 3 deploys an unsupervised document clustering framework to identify subject specialties in VABB-SHW, the Flemish academic bibliographic database for the social sciences and humanities. I study the disciplinary diversity of research specialties and the cognitive openness of disciplines to sharing subjects with other disciplines.

In the discussion I will reflect on the findings presented in this thesis and go through the main limitations. An agenda for future research is presented. We conclude by reflecting on the organization of the scientific system and evaluation practices considering interdisciplinary research.

Part 1 - Disciplines and interdisciplinarity

2 Disciplines

In this section I will sketch a brief history of the emergence of the modern scientific system, in which disciplines have served as structuring units. Starting in the 19th century, we discuss how changes in the ways of thinking about and organizing the sciences has instigated the structuring of academic system around the disciplines. We describe their main functions, namely providing a coherent cognitive framework, a common reference point for a tightly knit scientific community, and an organizational structure for education. The system of disciplines can be considered as a network of communications or an 'interactional field' linking and integrating the research system and the educational system with the outside (extra-academic) world by providing a common 'address' or reference point (Stichweh, 1992).

The high-paced growth and differentiation of the scientific system has brought about a decoupling of the traditional disciplines from their respective communities. As such, in line with Peter Weingart (2010) and Rudolf Stichweh, I argue that disciplines serve a different and more abstract function nowadays (mainly as a referral address to actors outside of the academic enterprise and as an administrative component grouping different research communities and specialties). The implications of this change in function are considerable for bibliometric research. Should they indeed be considered as central as they are for bibliometric analyses nowadays? Should we rely on existing disciplinary classification systems as a scaffolding for evaluation procedures? Do they adequately capture knowledge communities? I argue that the ongoing differentiation and hybridization of the scientific system imposes challenges to bibliometric research, which demands for new ways of approaching the cognitive system of the sciences.

2.1 A short history

From a sociology and history of science perspective, Rudolf Stichweh has extensively studied the coming into existence of disciplines as well as the functions they (have) serve(d) for the development and stabilization of the scientific system as we have known it over the past two centuries (Stichweh, 1992, 2003). According to Stichweh, disciplines emerged partially due to a change in thinking about scientific knowledge, and the social organization of the sciences. A first historical fingerprint of this shift is observable in the emergence of new structures in encyclopedias around the turn of the 19th century. From the Renaissance on up until that time, encyclopedias were made up of alphabetical listings of short paragraphs and articles describing discoveries, materials, or other empirical observations.

At the beginning of the 19th century, from being mere listings of scientific facts and discoveries, encyclopedias gradually changed into reflexive instruments. They became 'institutions for observing science'. 'Simultaneously', Stichweh writes, 'there was an increase in the use of organic metaphors to describe specific sciences and the connections among them'[11]. The observational stance taken in encyclopedias turned into a reflexive one and made thinking of the sciences as an organism or 'system' - independent of encyclopedias – possible. By trying to bring order into the information represented in encyclopedias based on the parts or subjects of the physical world they were attempting to explain, disciplines were gradually being used as representations of 'real systems' (Stichweh, 2003, p. 7; Yeo, 1991).

Around the same time, universities in Germany were reformed. The integrated universities from the past, which were centered around teaching and a central and hierarchically superior 'higher faculty' (medicine, law, and theology), gradually changed into a collection of separate research and education entities structured horizontally around the disciplines (the Humboldtian model of higher education called after the German minister of education at the time, Wilhelm von Humboldt) (Weingart, 2010). This organizational restructuring of the sciences further materialized the disciplines as entities structuring not only the cognitive content of the sciences, but also the social and professional life of academics (Abma, 2011, p. 26).

Communities of researchers stabilized around the disciplines in the form of scientific associations as well as organizational units in universities, pulling away scientific labor from the academies. This was a shift from the rather informal and independent communities which existed earlier. While universities in Germany were the first

[11] See for example the introduction of the term sociology by August Comte to describe his new diagnostic model for studying society. Earlier he had used the term 'social physics', establishing a strong connection with positivist and empirical sciences (see Heilbron, 1990).

ones to inherit this restructuring of both higher education and research around the disciplines, others in Western Europe followed soon after.

Disciplines from that period onwards thus developed both a cognitive and social identity or 'function'. This cognitive identity, which according to Stichweh (2003) precedes "the theories and methods with which disciplines work" (p 8), was already present before the emergence of the modern disciplinary system. Be it as rather fractionalized knowledge classified according to different materialities (or 'spheres') of the physical world. This shifted to a classification which was based on the knowledge system and research questions being pursued by the communities surrounding the disciplines. "From this point on, disciplines can be defined by their guiding research questions rather than by subject areas" (ibid.). While the re-classification of knowledge in encyclopedias according to disciplinary denominators facilitated the formation of disciplinary scientific communities, the organizational restructuring of higher education at universities further stabilized these communities.

As a social system that works with and modifies its cognitive content (or identity) disciplines can be thought of as heterogeneous networks of communications (Stichweh, 1992). Both these networks of communications, as integrated systems, and their constituting communications are linked to other communications. This linking is a process which constantly modifies itself, from one event to another. An important part of this network of communications are the scientific communities. These are groups of scholars specializing in the same subject matter, pursuing the similar research questions with tools and concepts from a common 'tool box'. Disciplines became a social system in which all scientific communication was organized.

2.2 A dual identity

In this section I will try to sketch the defining framework which I also use as a reference point throughout the remainder of this thesis. The description which I propose is a loose one and builds upon the existing sociological and bibliometric literature on conceptual aspects of disciplines and their operationalization. As I have set out above while briefly guiding the reader through the historical emergence of disciplines, the institutional component (the integration of teaching and research) of disciplines is one of the stabilizing aspects making disciplines stand the test of time (Stichweh, 1992, 2003).

For disciplines to exist and achieve continuity which outlasts the lives of separate research questions, problems, paradigms or even individual scholars, they need to be vested within an institutional framework along which staff hiring and undergraduate education in academia is organized. Disciplines as social institutions are also stabilized by their strong links to society. As Stichweh (1992) writes, by fulfilling specific professional needs in the economy at large, disciplines continue to be referral addresses for employee recruitment and other knowledge demands.

Jacobs (2013) attributes great importance to the social aspect of disciplines and uses the term institutional disciplines: "An institutional discipline is a recognized area of study that typically is identified with an academic department and an undergraduate major. … Without a department, there is no hiring, no stable employment, and relatively little faculty input into decision making" (p. 28-38). Jacobs is no exception in this regard; authors like Richard Whitley (2000), Stephen Turner (2000) and Andrew Abbott (2001) take similar stances in their work on disciplines. Hiring based on PhD diplomas obtained in the same discipline makes the structure re-create itself (Turner, 2000; Abbott, 2001, p. 127). According to these views, disciplines should mainly be seen as social systems (Hammarfelt, 2019).

While this institutional characterization gives the impression that disciplines can be demarcated along departmental lines and that the linear 'disciplining' and hiring processes put clear boundaries to each discipline, the system of today is dynamic: disciplinary boundaries in the modern scientific system are ever changing (see Stichweh, 2003). These dynamics of the disciplinary system follow from the cognitive work carried out by communities which are active within a discipline.

Disciplinary communities center around research questions and problems, and not so much around delineated specific subject matter as was the case in the 18th and early 19th century Europe. Interaction is often needed between the disciplines to come up with satisfactory answers to their questions (this is also exemplified by numerous scientometric studies demonstrating the 'web of science').

In addition, relationship between knowledge production in disicplines has moved ever closer to political, economic and social problems (Weingart, 2000)

Problem areas or knowledge domains treated by the scientific disciplines in that sense have become more fluid, especially when taking into account everlasting growth and internal differentiation of the sciences, which leads to increased numbers of disciplines and thus potential interaction between them (Weingart, 2003). The metaphor of oceans of knowledge proposed by Manathunga and Brew (2012) for the cognitive flexibility of today's disciplines in this regard is useful:

> "By and large, an ocean is constantly moving. Seen as an ocean, knowledge is wild, vast, unpredictable, treacherous, deep, windy, becalming, life-giving, fluid, liquid, powerful, invigorating. It has slipstreams, currents, waves and travel routes. New research specializations that emerge and then form part of the larger whole flow into it like rivers. Academics both individually and collectively bring together disciplinary spaces converging, merging, changing and challenging previous structures." (Manathunga and Brew, 2012, p. 53)

In *Undisciplining Knowledge*, Harvey Graff (2015) presents excellent examples of interdisciplinary contacts between disciplines spurring the development of new fields and interdisciplines, ranging from the humanities and social sciences to natural sciences. For Stichweh (1992) interdisciplinarity always happens. The mere existence of an interactional system of disciplines per definition leads to interdisciplinarity. Thus, while it is often assumed that the cognitive content created by a disciplinary community or with which a discipline works – is coherent and cohesive this is expected to be hardly the case.

To summarize, disciplines have a double identity: a socio-institutional and a cognitive one. The socio-institutional identity makes disciplines persist and stable. They serve as a referral address (Stichweh, 1992; 2002). The core of the cognitive identity, the central research questions pursued by a discipline and paradigms shaping puzzle solving endeavors within them, serves a similar function but is much less constant or 'fixed' in time. The very essence of scientific disciplines in obtaining legitimacy, both within the scientific system and society at large, is to innovate and communicate, leading to cognitive boundary crossing and interdisciplinarity by default (Stichweh, 2002):

> "There is normal science in a Kuhnian sense, always involved with problems to which solutions seem to be at hand in the disciplinary tradition itself, but normal science is always expanded upon by a parallel level of interdisciplinary science which arises from conflicts, provocations, and stimulations by other disciplines and their intellectual careers"

As we will see in the following section, the internal differentiation of the disciplinary system over the past century has reconfigured their boundaries in many ways. Departmental structures do not neatly overlap with knowledge communities (for example in the form of research centers), and teaching is sometimes being coordinated by different departments. The function of conducting scientific research is also diffused to other organizations than the university (trans-disciplinarity). We may ask if the outlook of disciplines and their integrated identity has shifted to a state which calls for other approaches than applying straightforward disciplinary classification models.

2.3 Differentiation

The development of the disciplinary system in the 19th century can be understood as a response to increased availability of empirical information and methodological innovation. According to both Stichweh and Weingart, communicative overload together with a growing amount of available resources (botch in terms of human and financial capital) brought about this important change in the social and classificatory organization. The internal differentiation of the disciplinary system which followed immediately after its establishment and is still ongoing today is also a consequence of its growth (Weingart, 2003). The rising number of specialties and disciplines leads to growth and intensification of interactions between these disciplines, a growth in interdisciplinary configurations (Stichweh, 2003, p. 6). Differentiation or specialization of the disciplinary system proceeds along different directions.

A first form being along disciplinary lines. The growth of the scientific system as we have witnessed not only made the disciplinary system larger, but also more fluid and versatile leading to the (continued) establishment of many sub-disciplinary and specialty groups working within disciplinary frameworks. For illustration purposes of such internal differentiation, we might take a look at the American Psychological Association. Mainly for reasons of convenience, because the history of association is well documented online[12]. Established in 1892, the organization counted 31 member and two organizational units. In 1930 and 1945 this number rose to 530 and 4,183 respectively.

While this is in itself spectacular, in 1970 the number further increased to 88,500. Today the association counts more than 121,000 members. A system growing this fast can either be restructured from the bottom-up, or further specialize (Weingart, 2003). The latter seems to be the case for the APA. After WWII, an influx of resources allowed for an expansion of – and the establishment of new universities in the US, attracting more research staff and thus enabling research into an ever-growing number of subjects. special interest groups were introduced (19 in 1944). From 1960 to 2007 34 more divisions were added. Many of these divisions are now established specialties, with their own set of core journals, conferences, graduate and PhD degree granting programs (e.g. sports psychology, health psychology, organizational psychology, etc.).

The development of these specialties or subdisciplines is also dependent upon change in society at large. In Belgium, clinical psychology got introduced as a graduate program in 1965 as a response to real societal needs (Richelle, Janssen, & Bredart, 1992). 'Under government sponsorship, psychiatric centers, mental health services, rehabilitation centers, family therapy centers and, last but not least, psychiatric wards in general hospitals were established' (ibid., p. 519). Differentiation of, in this case, the welfare and health system external to the scientific system has the potential to further propel or initiate disciplinary and specialty growth.

Some of these specialties show little overlap with neighboring disciplines or fields and function at the core of a disciplinary denominator (e.g. experimental or general psychology to stay with the example of psychology). These can be thought of as disciplined

[12] https://www.apa.org/about/apa/archives/apa-history

specialties. Others however function at the boundaries of a discipline. Clinical neuropsychology for example. According to the APA, the specialty ‘promotes interdisciplinary interaction among various interest areas including physiological cognitive developmental clinical rehabilitation, school, forensic and health psychology’. Their research groups or graduate programs are shared efforts between different disciplinary departments. Courses are thought by experts who are affiliated not only to the psychology department. They do have their own specialized journals and conferences, but a discipline like departmental structure is not in place. In a sense, they can be termed fractured disciplinary specialties (e.g. brain science and cognitive psychology).

Parallel to this a second form of internal disciplinary differentiation takes place; specialties without linear disciplinary predecessors are being established. This is often in response to societal or political needs (or external differentiation). If these questions developed in, say, politics are complex and ongoing (e.g. the monitoring of the outputs produced by the academic community), these specialty communities have the potential to develop into large scale discipline like communities of scholars.

A case in point of such a specialty is scientometrics. Scientometrics behaves much like a discipline. It has a well-developed cognitive identity, with a structured research program, devoted journals and conferences, but without an equally strong socio-institutional identity like, say, economics or communication sciences. Scientometrics as a field, being fairly large and established, is in some cases considered a subfield of computer science or economics, and in others of communication sciences but has no disciplinary history in the strict sense. In Flanders for example, the primary organization which undertakes scientometric research is ECOOM, the Centre for Research and Development Monitoring. Although ECOOM is an interuniversity consortium, it does not offer education, nor does it grant undergraduate or graduate degrees. Its primary mission is to assist the Flemish government in monitoring R&D efforts in Flanders. The ECOOM branches are affiliated to different departments at the different universities. And also, within the scientometric community differentiation again takes place. As the community grows, coherent groups of researchers are being formed around different subjects (i.e. innovation management, altmetrics, etc.).

But not only science itself is being ‘scientisized’, other aspects or systems of social life are also taken into the laboratory without having a pre-established disciplinary community by which they can be studied. Other examples of such interdisciplinary specialties are the ‘studies’ fields which in many cases have now become established programs, both in education and research. Gender studies at Ghent University, to take one example, is a shared effort between Faculty of Arts and Philosophy and the Faculty of Political and social sciences. The Centre for Research on Culture and Gender at Ghent University at the same time brings ‘together scholars and students across disciplinary divides’ and they ‘are also open to interdisciplinary collaboration and dialogue across the faculties, scientific paradigms and fields.’ https://www.crcg.ugent.be/en/about-crcg/.

These new emerging fields are per definition interdisciplinary, but without being ‘disciplined’ previously. In a sense, they are cross-disciplinary and fractured, they have multiple disciplinary communities who pay attention to the subject, without an integrated institutional structure attached (e.g. scientometrics, gender studies).

Some disciplinary associations have followed a trajectory, experiencing tremendous growth and subsequently specializing into sub-disciplinary units. Whereas communities of scholars were tightly knit around disciplines when these structures got first introduced in the 19th century, the growth of this system now redirects this function of ‘communities of communication’ to specialized units. They are replacing the discipline as a primary community of colleagues to which a researcher addresses her new findings and published research by establishing devoted associations or entirely new discipline-like structures on the sub-disciplinary or specialty level.

From communities of scholars centered around the discipline-based departments, quite soon after the installation of the disciplines as the primary referents both on the social as well as the cognitive level, differentiation made cognitive communities drift away from disciplinary cores or ‘identities’ and self-organize into smaller units of communication. The specialist communities are in many cases organized within or tightly associated with disciplines, but in other cases exist almost entirely independent of a disciplinary or academic core or ‘identity’ (without devoted departments, under-graduate and graduate curricula). Differentiation of the disciplinary system is in that sense changing the cognitive identity of the sciences, which demands for new ways of approaching this hybrid reality of disciplines and specialization. In this context of increased differentiation, the question becomes relevant to what extent we should oppose disciplinarity and interdisciplinarity, as they seem to co-exist. A more straightforward and enduring question about the unity and diversity of science is at stake here.

The expansion and specialization of the scientific system thus leads to changes in both its internal and external demarcations (Weingart, 2003). On the one hand, disciplinary identities are pluralized. The traditional disciplinary departments from the past, with a relatively small and specialized community established around a common denominator both on the social and cognitive level, have changed into heterogeneous collections of specialisms. Some of these specialisms have emerged or further expanded due to societal or political impulses. According to Weingart (2003; 2016), due to science's relative success in attaining legitimacy as a knowledge producing system and its further expansion, the primary function of the system (in terms of producing new knowledge and problem-solving techniques) is also diffusing to other systems in society. Concrete evidence for this process is the installation of technology transfer offices at universities, the growing number of research performing spin-offs and organizations, and the growing reliance on knowledge workers and research professionals in different sectors of society.

What used to be a rather clearly delineated knowledge producing system is to an increasing extent becoming a part of a broader knowledge society. For the development and emancipation of the criminological discipline in Belgium, the development of a crime prevention policy in the 1980s has been an important catalyst (Pauwels & Verhage, 2019). Today, next to and in collaboration with universities, criminological research is also conducted outside of the universities. The National Institute for Criminology and Criminalistics also carries out a great deal of scientific research. It maintains databases for research and publishes articles in national and international academic journals.

Thinking about science in terms of traditional disciplines and their identities might therefore not be that relevant anymore, in the future it "may be necessary to take the interaction of 'knowledge systems' into view.” (Weingart, 2003, p. 197). As new specialties and communities of scholars are being formed around subjects which have been largely unknown to traditional disciplinary communities, it becomes irrelevant to consider disciplinary identities by thinking about what is and what is not belonging to one or the other discipline (Massey, 1999). Instead, the disciplinary identity can be approached in a relational manner by studying disciplinary attention payed to subjects, or the disciplinary identity of subjects themselves (see Van Praag & Daenekindt, 2021) next to specialty formation. In addition, trans-disciplinarity, or the co-creation of new knowledge and problem-solving techniques with stakeholders outside of the university or academic system (e.g. criminology), is becoming an ever more important aspect when thinking about differentiation in and outside of the scientific system. But this of course would serve as a whole new topic for a PhD dissertation.

2.4 Scientometrics and inter-disciplinary identities

It is generally agreed upon amongst scientometricians that there is a lack of reflection on what the concept of disciplines entails (see Hammarfelt, 2019; Sugimoto & Weingart, 2015). I have tried to show that outside of the field of bibliometrics (sociology and history of science more specifically), there exists a nuanced and systematic exploration of how disciplines emerged, how they have differentiated and changed in form. And how, in some cases, they have become partially detached from the specialist communities of researchers.

For most scientometric studies of (inter-)disciplinarity or discipline specific research practices, however, catch-all and universalist classification systems are used. This way of working still gives high priority to the idea that the scientific system is rigidly organized along stable disciplinary lines and that these categories represent coherent specialist communities. While we have seen that this was initially the case, growth and differentiation of the system both along and across disciplinary boundaries has made the contemporary outlook more complex than what is represented in pre-defined classification schemes. Ongoing specialization and sub-disciplinary differentiation are not well represented by these sets of labels, or captured only after developing a discipline-like and recognized scholarly community. In addition, disciplinary or specialty development are context dependent (see the previous examples of the emergence of psychology and criminology in Belgium).

This is perhaps also one of the reasons why so much discrepancies in classifications of research output exist (Sīle et al., 2021), why cognitive and organizational classifications of scientific output do not always align well (Guns et al., 2018), or why different ways of measuring interdisciplinarity give different results (Wang & Schneider, 2019), to cite only a few problems. In essence, dominant scientometric operationalizations of disciplines presume a stable entity, or a collection of stable entities fitting into a classificatory straight-jacket. While this might certainly be the case for the socio-institutional identity of disciplines to some extent, their cognitive identity develops fast and is highly dynamic, with interdisciplinary interactions being the default rather than the exception.

3 Assessing Interdisciplinary Research in the Social Sciences: Are we on the right track?

Full reference: Eykens, J. (2022). Assessing Interdisciplinary Research in the Social Sciences: Are we on the right track?. In Engels, T. C. E., & Kulczycki, E. (Eds.). Handbook on Research Assessment in the Social Sciences. Edward Elgar Publishing. ISBN: 978 1 80037 254 2.

Interdisciplinary research or IDR is regaining a lot of attention in science studies and higher education policy alike (Tarrant & Thiele, 2017). The new adage in research policy making particularly seems to become: "Interdisciplinarity is here to solve the most complex societal problems". Assessment of scholars and research incorporates this idea and different indicators and assessment procedures are being developed to be able to better account for the "interdisciplinarity" of research output, research and/or teams. Large research funding organizations, like the European Commission and the European Research Council (Allmendinger, 2015), the National Science Foundation (NSF) in the USA, and the National Natural Science Foundation of China (NSFC) all share this idea and give high priority to interdisciplinarity, for example by developing specific financing opportunities for interdisciplinary research (IDR).

Reasoning goes that interdisciplinarity should be encouraged because it is bound to entail innovative and concrete, problem-oriented research. Interdisciplinary research also allows the boundaries of the various disciplines to be crossed. These assumptions are reflected in the definition of interdisciplinary research put forward by the National Academies (2004), which is common in policy documents, bibliometrics and research evaluation alike:

> "*Interdisciplinary research is a mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or*

> *more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice.*"

Fundamental breakthroughs are expected to emerge from interdisciplinary research; by means of the integration of knowledge from separate disciplines, one can face so-called "grand challenges" for which the solutions lie outside the boundaries of a single discipline. It is often assumed that integration cannot arise within the individual disciplines as such: new solutions and insights related to pertinent societal challenges are the result of the integration of information, data, techniques, tools, perspectives and theories or concepts from the different disciplines.

As Bonnie Wolff-Boenisch, head of research affairs at Science Europe stated, however, "interdisciplinarity is not new, but [in Europe] it has gained increasing traction in the context of the global transformation of societies, the SDG's, and the 'Mission-oriented research' concept of the European Commission" (Science Europe, 2019). Bibliometricians and science policy makers alike are working hard on the development of new indicators to measure different facets of IDR (Wang & Schneider, 2019). In the following, we will first highlight the main paths that are being explored in this race for indicators. We discuss the suitability of the approaches in light of the changing nature of disciplines and the social sciences as a whole. In a second part we discuss research into the qualitative strand of research assessment (e.g. peer review) in the context of IDR. In a third section we discuss seven principles put forward by Klein (2008) to guide interdisciplinary research assessment and peer review.

3.1 Bibliometric indicators for capturing interdisciplinary research

Although complex in nature, different bibliometric approaches have already been proposed for the identification and measurement of IDR. When proposing these indicators, the aforementioned definition of the National Academies is quoted almost by default. Strong emphasis is hereby placed on the need for a proxy that captures the ‘integrative component’ of IDR. Knowledge integration, it is argued, takes place when combinations of discipline-specific information arise. Those integrative combinations then, if sustained and continuous, ensure that the boundaries of the ‘disciplinary silos’ are transcended.

Certain bibliographic units (e.g., documents, authors, departments) are categorized ex ante on, for example, the basis of a discipline code list or the organizational affiliation of the authors is considered. It is than approximated whether - and to what extent - the integration of information from the various disciplinary units takes place on the level of disintegrated units (authors, publications, journals, research units) (Abramo et al., 2012; Abramo et al., 2018; Schummer, 2004). The relational and dynamic nature of interdisciplinarity and the role it plays on the discipline or field level is not fully captured. Next, we will discuss several examples of indicators for IDR and highlight which ones are used in research assessment exercises. Note that these indicators are not specific to the assessment of social sciences. As we will see further on, some caveats should be considered when applying these indicators to social science disciplines or fields.

3.1.1 Co-classification

Early indicators of inter- and multidisciplinary research include accounting for the relative number of co-occurrences of subject classification headings, and the mapping of co-classification relations (Tijssen, 1992). For a set of publications, co-occurrences of subject headings are counted. The premise behind this method: the co-occurrence of subject headings from different disciplines indicates if a publication or journal is of a cross-disciplinary nature. The ratio of out-of-field citations is another established approach (Amir, 1985; Tomov & Mutafov, 1996). Here the number of citations from a publication or journal to journals from other disciplines or fields relative to the number of citations to journals from the same discipline are considered. Following this logic, Amir (1985) and Abramo et al. (2012) also study the disciplinary affiliation of authors. The idea is straightforward: if the co-authors of a scientific publication originate from different disciplines, the research can be considered as multi-or interdisciplinary.

3.1.2 Network centrality

Another popular indicator on the level of journals and their citation environments has been proposed by Leydesdorff (2007). The betweenness centrality (or intermediation), a measure borrowed from social network analysis is applied to estimate whether a journal acts as a knowledge broker within the journal-journal citation network. The betweenness centrality gives an idea of how many times a journal is located on the shortest path between other journals (Leydesdorff, 2007; Leydesdorff & Goldstone, 2013).

3.1.3 Diversity based indicators

More recently, indicators have been proposed that take into account not only the degree to which work from other fields is cited or whether a unit is classified under multiple subject headings (Porter & Rafols, 2009). Measures of diversity, such as the Rao-Stirling diversity (Porter & Rafols, 2009), lead to indicators of interdisciplinarity which also take into account the balance of the distribution of different disciplinary categories in the reference list, and the disparity between these categories. For these more complex indicators, just like the ones discussed earlier, the references of a publication (or all publications in a journal) are classified on the basis of a predetermined discipline classification scheme (note that this is only partially true for the centrality indicator). The diversity indicators have been applied by the ERC in its "comparative scientometric assessment" of the results of the ERC funded projects (Macaluso, Pollitt, Gunashekar, & Larivière, 2015). However, research into the consistency of the results emerging from the reference-based diversity indicators points to fundamental inconsistencies (Wang & Schneider, 2019). The complexity and versatility of interdisciplinarity is cited as a possible explanation for divergence of the results.

3.1.4 Text based approaches

Text and topic analysis are also used to develop indicators for IDR. The same premise also applies to this research strand: integration of discipline-specific information, now in the form of textual data, from two or more disciplines equals interdisciplinarity. The text-based indicators have received less attention in the literature so far, but, as we will show, have significant similarities with co-classification and citation and reference-based approaches. Text-based approaches are also used for indicators of IDR with the authors. An indicator used at the NSF to capture IDR is based on the identification of co-funding arrangements. The rationale is the following: if different directorates (which are discipline-based) fund a single award, then it can be concluded that we are dealing with IDR. We should thus be able to identify and subsequently measure IDR by calculating the percentage of awards contributed by each directorate.

Nichols (2014) developed an indicator which is now also being used at the NSF (USA). Nichols' model works with disciplinary topic bins. Topics are extracted from project proposals using text-mining techniques (LDA) and assigned to disciplinary categories. As an indicator for the interdisciplinarity of a document, the number of disciplines that are linked to the topics is counted. Then a diversity index is calculated to consider aspects such as variety and balance of different disciplines contributing to the project proposal.

Evans (2016) developed a similar corpora-based approach. Evans uses the corpus (the raw text from publications) for each disciplinary category in the Web of Science Subject Categories. The corpora of the authors are classified, and on the basis of similarity scores it is then calculated how similar or different the corpus of the individual researcher is compared to the disciplinary corpora. The IDR indicator for an author then starts from those similarity scores between the author's corpus and that of the four closest disciplines.

Gowanlock and Gazan (2013) take a slightly different approach, although, like Evans, they also propose an indicator at the level of the individual authors. To this end, they work with cluster membership of the publications of authors active in astrobiology. A cognitive map (using text clustering) of the publications is first drawn up on the basis of titles and abstracts. As a result of this clustering, the authors find a number of prominent subdomains. The interdisciplinarity of individual researchers is then calculated on the basis of the number of publications an author has in the different subdomains, as shown by the clustering; '[A]n author with publications in many clusters indicates they are engaged in interdisciplinary research, or perhaps they are not, but should be' (ibid., p. 158).

3.2 Processes of growth and change in disciplinary structure

Current evaluation measures and indicators designed for capturing IDR have their flaws (Vugteveen, Lenders, & Van den Besselaar, 2014; Wang & Schneider, 2019), yielding inconsistent results when positioned next to each other. Discussing the reasons why these indicators might be problematic Vugteveen et al. (2014) state that '[t]he diversity and coherence indicators depend on boundaries between fields. … This implies that the dynamics of the disciplinary system is not taken into account when calculating diversity and coherence.' On a fundamental/conceptual level, it is not clear what these indicators measure. They are designed to capture degrees of knowledge integration (Porter & Rafols, 2009; Wagner et al., 2011), but by simply looking at different aspects of reference lists change is not taken into account, it is not entirely clear if it is cross-disciplinary or interdisciplinary what is being gauged. The actual integration of knowledge takes place within the text and can yield new interdisciplinary configurations on the discipline or field level. Ultimately, it is this transformative aspect, the processes of disciplinary change that arise from knowledge generation and not the outcome of referencing behaviour as such, that makes interdisciplinarity unique.

As we will try to show, the fluidity of boundaries and heterogeneity of the fields evaluated becomes an issue. For the social sciences this can be extended. Just like the knowledge subjects studied in the social sciences, the fields themselves are more fluid and dynamic than those in STEM fields. Vugteveen et al. (2014) contend that a bottom-up approach, which takes into account disciplinary stability and change is required in this regard.

On the level of individual disciplines, differences between fields make matters more complex. First, there is a difference in knowledge production practices. Some fields are more inclined to journal publishing than others, naturally making them more 'open' or 'closed' to other fields that are more or less inclined to journal publishing. If in field A more articles are produced than in field B, changes are lower that ties exist between A and B, than between A and C. If C also produces more articles. Or source material for that matter. Disciplines which are more locally oriented will perhaps not be covered adequately when using commercial databases for research evaluation purposes.

Epistemological differences exist as well. The matter or subject which is being studied automatically makes some fields more or less susceptible to interdisciplinarity. Professional fields like educational sciences and law will be less susceptible to interdisciplinarity than fields in which theorizing about social phenomena plays a big role (i.e., sociology). Social sciences scholars often contribute to policy research, professional guidelines, etc. Which are not accounted for in commercial databases. If IDR leads to more societal impact and problem solving we might be missing out on important research output when we make use of current bibliometric indicators.

3.3 Disciplinary dynamics perspective

The aforementioned approaches to IDR indicators, both on the basis of references and on the basis of textual data, place a strong emphasis on knowledge integration as continuous (unique) knowledge flows - in the form of stable relationships between different units and their diversity. These methods are very popular and indeed make it possible to estimate to what extent a researcher (or a publication) uses information from different fields in her research at a certain point in time. However, in the context of IDR, these indicators take little account of the changing nature and dynamics inherent in scientific research. Scientific fields in the social sciences, as discussed by Bonacorssi (2022), are in constant flux, and may be more multidisciplinary at a certain point than later in time. The intensity of the relationships that specialisms have with other fields can also vary greatly over time (van den Besselaar, 2019; van den Besselaar & Heimeriks, 2006; van den Besselaar & Leydesdorff, 1996).

van den Besselaar and Heimeriks (2001) and van den Besselaar (2019) therefore propose an alternative perspective, drawing attention to patterns of change within the scientific system. According to van den Besselaar (2019), our bibliometric knowledge of interdisciplinarity and how it evolves over time is still too limited to develop adequate indicators. He proposes that we should first approach interdisciplinarity as one of the many components of the continuous processes of change of scientific disciplines. The changing position of journals in the journal landscape is central to van den Besselaar's work. The author cites an important argument for this, one that has also been raised several times in other studies (Abbott, 2001b; Jacobs, 2013; Stichweh, 1992): interdisciplinarity is not a characteristic, but a phase in which disciplines can find themselves at a certain point in time.

The role of interdisciplinarity in the social sciences

Practically all of the disciplines that are now considered as being established units of science had an interdisciplinary start, and may later play a role in the establishment of interdisciplines. Knowledge continuously flows from one discipline to another (see e.g., Jacobs, 2013, among other parts chapter 6 in particular); the integration in terms of referring to work from other fields is therefore not very special or unique. Scientific specialties are, to a greater or lesser extent, constantly involved in interdisciplinarity. The stabilization of integration processes in the form of new interdisciplinary fields, is the kind of sustainable interdisciplinary integration we want to understand.

For example, sociology, as proposed by one of its founders (August Comte) at the beginning of the 19th century, was based on biological ideas. A number of key concepts that Comte used for his sociological theories came directly from biology, and "his idea of diagnosis followed a medical model" (Heilbron, 1990, p. 262). Harvey Graff (2015) discusses a more recent example with the emergence of communication sciences, a now established field of research that arose from cross-pollination of insights from sociology, political science, psychology, etc. (Graff, 2015, pp. 55-56; Leydesdorff & Probst, 2009).

According to this view, interdisciplinarity is better understood in terms of developmental phases, a temporary stage through which parts of the scientific system move at a specific point in time, whether as a new interdisciplinary field or as a catalyst in the emergence of a new domain. Over time, a new scientific community can flourish from interdisciplinarity, which may stabilize around a number of problems and develop into a 'normal' disciplinary constellation (Jacobs, 2013; Light & Adams, 2017; van den Besselaar, 2019).

The focus is thus shifted to dynamics - changes in the science landscape - instead of static representations of integration calculated on the basis of ex ante classifications, which we mostly fall back on today in bibliometrics. According to van den Besselaar (2019), when we shift attention to processes, we will be able to observe other changes in addition to interdisciplinarity and specialization; the emergence of new fields, the growth and decline - or differentiation (disciplines that split into sub-specialisms), integration of specialisms, divergence or convergence, and the extinction of certain domains.

van den Besselaar does not reject the central idea of integration. Integration is indeed an essential characteristic of interdisciplinarity, but the use of knowledge from other fields (as measured by the diversity of reference lists) actually reflects only to a limited extent whether knowledge is effectively integrated in an innovative way and where this will subsequently lead. This is a common criticism also cited by bibliometricians and is based on the fact that a researcher, journal or publication cites sources from two or more other disciplines, but that this does not necessarily mean that he, she or it is an (more or less) interdisciplinary researcher, journal or publication.

Attention to processual aspects of knowledge integration can provide insight into those uncertainties. In this sense, integration of knowledge from different disciplines is part of the development dynamics of scientific fields. A field integrates knowledge from two or more domains and therefore (to a certain extent) also resembles the fields it integrates. To gain insight into this, van den Besselaar analyses the citation environment and the position of journals. The author starts from a disciplinary core journal for which he determines the citation environment. He then applies a factor analysis. The factors consist of the journals that exhibit similar citation behaviour. Journals with the same citation behaviour will therefore together form one factor (Van den Besselaar, 2019, p. 3).

3.4 Peer review and the qualitative assessment of interdisciplinarity in the social sciences

As we have seen, indicators of IDR are 'booming'. However, studies of the qualitative evaluation are less numerous. In recent years there has been a slight increase in the number of articles and chapters that look into the social and cognitive dynamics present in (peer review) panels tasked with evaluating (ID) research proposals. The existing body of literature on this topic identifies 3 central caveats to consider when designing qualitative assessment procedures for interdisciplinary research in the social sciences. (i) The cognitive characteristics which are valued in research communities are different across (social science) disciplines and (ii) the socio-structural characteristics of social science disciplines differ. It follows that (iii) interdisciplinary endeavours yield unique configurations that require tailor-made evaluation procedures. In the following we will briefly clarify each of these caveats and point toward literature relevant for the context of the social sciences.

The cognitive characteristics valued in research are different across social science disciplines

Not all disciplines have the same standards for evaluating research, making it difficult to develop a common yardstick for assessing (interdisciplinary) research. Guetzkow, Lamont, and Mallard (2004) for example, point toward one of the main criteria put forward when reviewers assess research quality: originality. The authors study the importance of originality in the context of peer review panels from five multidisciplinary fellowship competitions in the US. Relying on interviews with panelists, the authors study how originality is conceived and defined across disciplines. Diversity is found in the ways panelists from the social sciences and the humanities define originality. According to their results, humanists privilege 'originality in approach', whereas social scientists mention originality in terms of method (Guetzkow et al., 2004, p. 190).

Again from the perspective of disciplines, well established methods are often in place. As pointed out by Bonaccorsi (2022), however, even within one discipline (e.g. sociology), different epistemic communities exist which prefer their own set of methods. While this is often common knowledge for peers from that disciplinary community, we cannot expect this to be the case in the context of interdisciplinary collaboration. In many cases of IDR, no established methodological framework is in place (see Bammer, 2016). It is often devised on the go, or it even is the very essence of the research project in the first place.

These are two simple yet important examples one should keep in mind when bringing together or heading a panel of experts tasked with evaluating interdisciplinary proposals or research outcomes.

The socio-structural characteristics of social science disciplines differ

The process of evaluation has shown to be deeply interactional and social (Lamont, 2009; Lamont & Guetzkow, 2016). Lamont and colleagues have conducted a series of in-depth studies on the 'informal' structure of multidisciplinary panels in the social sciences and humanities in the US. They show that members from different disciplines have their own ideas of what excellence means and how it should be assessed (cf. the 'cognitive' aspect of originality in research as discussed above). These ideas of excellence, though, are often very implicitly present. In addition, it is found that personal preferences of reviewers play an important role too.

The more formal communicative structure differs across social science disciplines. Bibliometric studies of the social sciences have pointed out that the members of some disciplines prefer to communicate their findings in book format, while in other fields it might be more common to publish journal articles (Kulczycki et al., 2018). The rate with which publications appear differs between the fields as well. Additionally, citation practices vary greatly between different social science disciplines and even subdisciplines (Larivière et al., 2006; Nederhof, 2006). Using a uniform set of indicators within a peer review context might therefore be a bad idea. Especially in the context of interdisciplinary research proposals, it is advisable to take into account these variabilities.

Interdisciplinary endeavours yield unique collaboration configurations

Interdisciplinary research or collaborations are always unique in that attempts are made to find new combinations, or integrate knowledge or techniques and methods so that researchers are able to counter specific societal or scientific problems. Whereas it has been shown that some disciplines are less resistant for their peers to perform interdisciplinary research (see for example Porter & Rossini, 1985), researchers from some social science disciplines, e.g. economics, have been shown to be more insular. Contrary to what is commonly held for true, the research conducted by Porter and Rossini (1985) suggests that reviewers typically rarely criticize cross-disciplinary features of proposals. While it is shown that researchers prefer work belonging to their own specialty, the aspect of interdisciplinarity or cross-disciplinarity of research proposals is celebrated as a quality.

In addition, when the knowledge or techniques that are combined originate from disciplines that are cognitively more similar to each other, the efforts to be made to promote integration can be less problematic. The transfer of epistemological premises and vocabularies is more straightforward when approaches from two cognitively proximate social science disciplines (e.g. social geography and sociology) are combined than when this is the case for two more distant fields (social geography and health sciences). This cognitive aspect which should be kept in mind is also transferable to the socio-structural. Referring to the latter example in which knowledge from social geography is integrated with insights from medical fields (say, epidemiology) it should be kept in mind that completely different publication practices and evaluation criteria are prevalent in both fields. Every interdisciplinary endeavor thus consists of a unique configuration which requires tailor made evaluative procedures.

3.5 Seven evaluation principles for assessing interdisciplinary research in the social sciences (Klein, 2008)

To tackle the difficulties which arise in an interdisciplinary research assessment context, seven evaluation principles are proposed by Klein (2008) in her review on the subject. These principles are a bundling of many years of experience with inter- and transdisciplinary research studies and policy-making on the part of Klein, but also on the part of research management and policy systems currently in place which can be considered good practices. Here we recapitulate these seven principles in the context of IDR in the social sciences, and briefly discuss each one. For a detailed discussion of these principles, we refer the reader to Klein's original article (2008). While every project proposal or research outcome is indeed unique, these seven generic principles can serve an important function when designing evaluation or research assessment procedures for the social sciences.

(1) *Variability of goals:*

To begin with, not all fields or disciplines in the social sciences harness the same goals. It follows that the individual researchers from these different disciplines will behave differently. Whereas scholars active in more traditional disciplines might have the ambition to create new knowledge about a topic central to their field, researchers from subfields like feminist studies or area studies might have the ambition to empower certain groups of people. The same holds true for interdisciplinary research projects. For some, "the production of new and broad knowledge of a particular phenomenon" is important, and for others "the development of technical equipment or products" is the main goal (Klein, 2008).

(2) *Variability of criteria and indicators*

The previous principle "drives the variability of criteria and indicators" (Klein, 2008). More traditional indicators, such as the number of publications or citations, for example, are not equally applicable to all disciplines in the same style. When it comes to communicating research, some social science disciplines or specialties value publications in journals more, while other value books as outputs. The same goes for interdisciplinary research. While some projects might be concerned with societal changes, others will be directed towards the development of new scientific methods or techniques to approach a research problem. It goes without saying that these socio-structural differences as well as the differences in perceived goals should be taken seriously by panel members when assessing project proposals and their submitters. Societal impact, for example, should not be assessed with bibliometric indicators only.

(3) *Leveraging integration*
Integration is considered to be central to interdisciplinarity. It is therefore crucial to take into account the degree to which initiatives are taken to accomplish or 'leverage' this goal. Klein cites the organization of structural support to allow for integration, like opportunities for communication (meetings among researchers), the development of a common vocabulary, etc. A set of guiding questions has been developed by Klein to take stock off this aspect (see: Klein, 2004).

(4) *Interactions of social and cognitive factors in collaboration*
Interdisciplinary research, like all research, is a social process. Leveraging 'intellectual integration' (the previous principle) is a social endeavour and, according to Klein and others, communication and negotiation 'lie at the heart' of this endeavour. To assess these complex social and cognitive factors, a guide has been provided in the context of evaluating and studying projects in European research institutes (Bergmann et al., 2005).

(5) *Management, leadership, and coaching*
Here again emphasis is placed on "how well the organizational structure fosters communication". Leadership is an important aspect in this regard, and should thus be taken in

consideration when an interdisciplinary research project entails complex collaborations among researchers from different (and disparate) disciplines.

(6) *Iteration and transparency in a comprehensive system*
According to Klein, a strictly linear evaluation model is not appropriate for the assessment of interdisciplinary research. IDR in many cases develops in different phases and reiterates over these phases. In an early phase, principles 4 and 5 might be very important and thus deserve more attention when intermittent assessments are carried out. In a later stage, when an IDR project comes to an end, indicators for research output or impact might become more important. Transparency ensures that evaluators and those being evaluated are aware of the criteria being used at what stage. Ideally, Klein suggests, both evaluators and those who are evaluated get involved when defining appropriate indicators for their goals.

(7) *Effectiveness and impact*
The principle of effectiveness and impact returns to the first two principles. The impact of IDR is often "diffused, delayed in time, and dispersed across different areas of study and patterns of citation practice" (Boix-Mansilla, 2006). Thus, it is required for the assessment of IDR to consider it thoroughly and ideally take into account possible but unpredictable long-term impacts.

Most of these principles require an active conversation among those who submit proposals and are conducting interdisciplinary research, and those who evaluate IDR. Appropriate evaluation, Klein states, is not given but made: "It evolves through a dialogue of conventional and expanded indicators of quality". As we discussed earlier, this is because 'peers' in the traditional sense are largely lacking in the case of interdisciplinarity. As such, "there is no consensus on the legitimate sources and types of control over it" (Huutoniemi & Rafols, 2017). A co-creation model of evaluation procedures similar to the model described by (Laudel, 2006) and discussed in the next section and guided by the principles listed above, might lead to more appropriate research assessment practices for IDR.

3.6 Discussion: implications for research assessment in the social sciences

Disciplines are important (if not, the most important) structuring components in modern academia in the sense that they serve as a cognitive address - with their curricular structures in the educational system (Stichweh, 1992) and their organizing structures for knowledge communication amongst its members, they define and constantly redefine the boundaries with other disciplines by putting forward expectations for its (new) members (i.e. what is the required knowledge for students to be allowed to join the ranks of the discipline, what are the appropriate methods to be used, what are the questions which need to be addressed, and what are the topics to be studied).

The latter is done by the 'gatekeepers' of the discipline (Becher & Trowler, 2001; Gieryn, 1983; Lamont, 2009). These gatekeepers are teachers and researchers who sit on review boards of evaluation panels that decide upon funding, the panel members or organizational committee members of conferences, reviewers or editors of journals, etc. For knowledge to be accepted by the disciplinary community, or for that matter, a project to be executed by members of a discipline, the manuscript or project proposal needs to pass the gatekeepers of that community. As we briefly touched upon in this chapter, all disciplines have a set of either explicit or implicit criteria and rules by which the gatekeepers assess new knowledge or knowledge in the making and make decisions about their approval. As discussed, quality criteria differ among disciplines. These differences depend on cognitive / epistemological characteristics and/or variations in publications practices.

Early-stage cross-disciplinary knowledge configurations, however, are often innovative collaborations between different disciplines, or forms of integration of knowledge originating from the boarders - or within different disciplines. As we have discussed, there is not one reference group, but two or more. New cross-disciplinary knowledge which is in the making, might be of relevance for more than one disciplinary community. There is no one single set of rules or criteria to evaluate if the new knowledge will be of benefit, or if the project meets the quality standards of all the different disciplines.

Quantitative indicators in that sense might be of some use to assess candidates with regard to their disciplinary peers, but in the case of interdisciplinary project proposals, the disciplinary heterogeneity of the candidate base makes it largely impossible to make use of simple quantitative indicators. If we wish to capture new and innovative interdisciplinary configurations, the usage of ex ante classifications can be problematic for the reasons discussed in this text. Science mapping might be used to get a bird's eye view of the scientific landscape and orient those tasked with evaluation towards knowledgeable peers or experts who are cognitively 'closer' to the researchers undertaking interdisciplinary adventures. Our knowledge of interdisciplinarity and disciplinary dynamics is hardly sufficient to allow for the development of appropriate indicators of interdisciplinary configurations, and this is especially the case for the social sciences. Here we briefly recapitulate what the purpose of science maps could be within the context of research evaluation and interdisciplinarity.

Science mapping

In line with van den Besselaar and colleagues we have argued that we should first approach the scientific system in terms of dynamics of change. With regard to quantitative approaches discussed above, a first step than consists of adequately mapping the scientific system. For the social sciences, however, citation-based approaches can be problematic. To summarize: publication practices in the social sciences differ from those in STEM fields. Whereas most of the original research produced in STEM fields appears in journals and conference proceedings, studies in social sciences often appear in book or book chapter format. Many insights get distributed in local journals or policy briefs and working papers as well. The latter are often inadequately covered by commercial reference indexing services. We have also touched upon epistemological differences between the social sciences and STEM fields (discussed by Bonaccorsi, 2022) as a complicating matter.

Science maps open up the possibility to study changes in the disciplinary system and will allow us to come up with more adequate and dynamic approaches to IDR. The increase in data availability (i.e. more textual data) will allow researchers to not only take into account journal article publications, but also other research outputs in the form of text when drawing these maps. Sidestepping the need for predefined science classifications, a bottom-up text based which makes use of document similarity methods and clustering for example, could yield important insight into the social sciences landscape. In an evaluation context, these methods allow research administrators or policy advisors to locate research or researchers on the boundaries of established fields and disciplines: the cognitive localities where knowledge integration takes place. Next, we briefly discuss the implications of what has been summarized in this chapter for qualitative assessments and peer review in the context of interdisciplinarity in the social sciences.

Qualitative assessments and peer review

Gritt Laudel (2006) describes peer review processes at two collaborative research networks in Germany (funded by Germany's most important funding agency for university research, the DFG). A network consists of about 10 to 20 research groups from different specialties. The funding programs aim to promote IDR. Although this study does not address social sciences as such, it shows in detail how co-creation of evaluation procedures can be of great value. A review setting is created in which applicants are consulted throughout the review process to ensure interdisciplinary learning of the reviewers. While the project is being developed, this way of organizing the review process allows for the formation of a 'project community' instead of disciplinary community. As Laudel states, however, the applicability of this procedure appears to be limited to areas where 'IDR is common, and where IDR is only moderate'.

The principles addressed by Klein (2008) and discussed in this chapter are less imperative in that they do not put forward a specific way of working, but they are all designed to facilitate a similar cross-disciplinary learning in which a community of scholars and reviewers is formed around a specific interdisciplinary knowledge making endeavour. And by doing so, researchers and reviewers might become more aware of each other's epistemological preferences and or disciplinary cultures. These principles for research assessment developed by Klein should be seen as an important first step towards more appropriate evaluation procedures for IDR.

3.7 Concluding remarks

Interdisciplinary research is lauded for its transformative qualities. Innovative IDR has the potential to change the scientific landscape and reconfigure disciplines or even lead to the emergence of entirely new fields. However, the statement that IDR leads to more qualitative problem solving than disciplinary research and is indeed capable of solving grand challenges merits further research to substantiate these claims. As Jacobs and others have pointed out, specialized disciplinary research has led to ground-breaking research already, and should not be considered minor when compared to IDR.

As we have seen, IDR has been around since the early establishment of the modern disciplinary system. Disciplines should not be taken for static and natural; they are social and dynamic entities which should be studied and approached as such. The disciplinary dynamics perspective, as has been introduced by van den Besselaar and which has been discussed in combination with science mapping in this chapter can be seen as an important first step in the bibliometric identification and approach of IDR and disciplinary change. For qualitative research assessment and peer review on the other hand, we propose the guiding principles introduced by Klein as important cornerstones.

In the first part of this chapter, we have presented a brief review of the state of art of bibliometric indicators that are being developed to assess and take stock of IDR. We argue that IDR is not so special or unique as one might expect from the descriptions and beliefs found in research policy documents. As stated by sociologists and science historians, IDR has played a significant role in the development of the sciences since the actual emergence and birth of the disciplines themselves. Therefore, IDR still plays a vital role in research policy and the development of the scientific system as we currently know it. Unfortunately, indicators of IDR reveal only part of the complex change processes taking place within the sciences. In this chapter we therefore argued that we need research into the dynamics of disciplinary growth and change before we can adequately decide what we want to measure and indicate about interdisciplinarity before we develop more applied indicators.

In the context of qualitative research assessment (e.g. peer review), we have briefly highlighted three central concerns that should be kept in mind when we design evaluation procedures for interdisciplinary research. First, it should be clear from the start that different disciplines and fields have different conceptions of research quality and excellence. This translates to cognitive aspects about what research originality is about, but also socio-structural aspects of the disciplines (cf. supra, section 3.1 and 3.2). What are the preferred communication formats, and what role do citations play in the individual fields.

We also emphasize "IDR creates a new boundary within the academy". As As Huutoniemi, Klein, Bruun, and Hukkinen (2010) state: "an operational definition of such research, plus a set of viable parameters to empirically distinguish it from disciplinary research – a problem that is not yet fully solved [...] The participation of researchers in the definition of criteria and the selection of reviewers ensures that more aspects of the work can be more comprehensively assessed. Such a dialogue and feedback loops between researchers and reviewers also supports a mutual commitment to long term goals" (Huutoniemi, 2010). We pointed out that "the participation of researchers in the definition of criteria and the selection of reviewers ensures that more aspects of the work could be more comprehensively assessed. This ongoing dialogue and feedback loops between researchers and reviewers also supports a mutual commitment to long term goals" (ibid., 2010, p. 313).

As Huutoniemi and Rafols (2017) conclude: "Although IDR has by definition many characteristics that make it particularly difficult to evaluate, it is important to note that there is also much contingency and variation within disciplinary research. Quality and performance are relative not only to disciplinary standards, but also to the goals, expectations, norms, and values of stakeholders and thus vary from one evaluation context to another". The seven principles for research evaluation proposed by Klein (2008) have therefore been reiterated as an important guideline.

Part 2 - Text-based approaches to science classification and clustering

4 Fine-grained classification of social science journal articles using textual data: First steps

Full reference: Eykens, J., Guns, R., & Engels, T. C. E. (2019). *Article level classification of publications in sociology: An experimental assessment of supervised machine learning approaches.* In G. Catalano, C. Daraio, M. Gregori, H. F. Moed, & G. Ruocco (Eds.), 17th international conference on scientometrics & informetrics (ISSI2019) (Vol. 1, pp. 738-743). Sapienza University of Rome, Italy: Edizioni Efesto.

Classifying scientific articles according to disciplines is most commonly done by making use of a proxy such as Clarivate Analytics' Web of Science (WoS) journal level Subject Categories (SC). Clarivate's staff working on WoS assigns the journals it indexes to one or more SCs and the publications that appear in these journals are treated as belonging to the same SC (for details on this procedure, see footnote 1 in Pudovkin and Garfield, 2002). Although this has proven to be useful, it has been treated as a limitation as well. Glänzel, Schubert, and Czerwon (1999), for example, point out that such an approach works well in the case of highly specialized journals, but that it is problematic for publications appearing in multidisciplinary or general journals. Solutions to this problem involve article level (re-)classifications based on the SCs of the references made in the article (cf. Glänzel et al., 1999), or clustering articles based on their citation relations (cf. Waltman and van Eck, 2012).

In the citation clustering study by Waltman and van Eck (2012) the authors distinguish between three different levels. At the highest level, the clusters correspond to 'broad scientific disciplines' (i.e. 'natural sciences', 'social sciences', etc.), and at the lowest level to 'small subfields' (i.e. 'library and information science'). This lowest level can be perceived of as research specialties, and consists of 22,412 clusters. Classifying publications on such levels of granularity is, as the authors acknowledge, difficult (Waltman and van Eck, 2012, p. 2386). Finding and manually assigning labels that adequately denote the clustered ‘communities’ on this scale becomes virtually impossible. Citation and referencing data are more often than not unavailable or only partially indexed by existing citation indices.

For publications in the social sciences and humanities (SSH) these concerns regarding citation approaches can be further extended. First, the lack of coverage in indexing services like WoS poses additional problems. Not all sources used in the SSH are covered by WoS or Scopus, which means that citation clustering will only yield a partial result. Second, when aiming for reference-based reclassification (see Glänzel et al., 1999), determining the SCs of the references of a publication is problematic, not only because of the lack of coverage of SSH journals, but also because SSH scholars tend to cite more books and non-source items (Ossenblok, Engels, and Sivertsen, 2012; Larivière, Archambault, Gingras, and Vignola-Gagné, 2006).

Making use of other sources or publication meta-data to classify publications, like for example author affiliations (cf. Guns, Sīle, Eykens, Verleysen, and Engels, 2018) might help to overcome these hurdles. But, as Guns et al. (2018) point out, author affiliations do not always correspond well with the cognitive domain SSH authors are working in. In times of increased specialization, moreover, departmental affiliations, like WoS SCs, are often too generic to get an adequate understanding of an author’s expertise.

In this article we aim to counter the previously mentioned shortcomings by applying a text-based approach. We present the experimental results of a supervised machine learning (ML) exercise and assess its potential for a fine-grained automated classification of scientific articles in sociology.

4.1 The Flemish Discipline Code List (DCL) (2018)

The Flemish Discipline Code List (DCL) is used as our guiding classification scheme (Vancauwenbergh and Poelmans, 2019a). The DCL is structured as a hierarchical tree with four levels. The first level refers to 7 broad fields of science. To allow for international comparison, the team working on the classification scheme has opted for conformity with the highest level of the OECD Fields of Science coding scheme (2007) (hereafter referred to as FOS). For the case of sociology and anthropology, on the top level of the FOS we find category 5 'social sciences' and a rather generic sub-category '5.4 Sociology and Anthropology'.

Figure 1. Excerpt of tree structure: OECD FOS (2007) coding scheme and DCL (2019).

<u>Hierarchical structure of OECD FOS (2007)</u>

5. Social Sciences
    5.4 Sociology and Anthropology

<u>Hierarchical structure of DCL (2019) (ex. "050405")</u>

05 Social Sciences
    0504 Sociology and Anthropology
        050401 Anthropology (14 subcategories)
        050402 Applied Sociology (11 subcategories)
        …
        050405 Social Change (2 subcategories)
            05040501 Social change
            05040502 Social movements and collective action
            05040599 Social change not elsewhere classified
            …

The third and fourth layer of the DCL add two more granular layers representing disciplinary subfields. The third layer might be interpreted as referring to sub-disciplinary categories, whilst the fourth level can be considered as referring to research specialties, in this case within sociology. To construct and define this most granular level, experts from the corresponding fields were consulted. In total, the DCL contains 2,866 codes (for details, cf. Vancauwenbergh and Poelmans, 2019a). Our objective is to automatically classify abstracts into the most granular categories of the coding scheme.

4.2 Methods

Since 30 years, the dominant paradigm of Text Classification (TC) consists of ML approaches. ML algorithms are deployed such that "a general inductive process automatically builds an automatic text classifier by learning, from a set of pre-classified documents, the characteristics of the categories [or labels] of interest" (Sebastiani, 2002, p. 2). ML approaches have already been applied to classify abstracts of journal articles (also full-text or parts thereof) (for a recent example, see Langlois, Nie, Thomas, Hong, and Pluye, 2018). Our approach is different from previous studies as we select articles and abstracts from one social sciences discipline, sociology, and assess the accuracy of supervised ML algorithms for classifying these abstracts into 77 different sub-disciplines, or specialties.

4.2.1 Data collection

As noted, the development of supervised ML models involves collecting a set of pre-classified documents; the training data. Based on these data, a model can be learned with which we can predict an output for a yet unclassified input. Put differently, the training data consists of examples of already classified input/output pairs. As we envisage a granular classification of abstracts in sociology, the Sociological Abstracts database (SA) was used to construct a training data set.

SA covers articles published in 1000+ journals in the field of sociology and related social sciences (i.e. anthropology, social psychology, etc.), dissertations, books, conference papers and proceedings dating back to 1952. The sources covered are relatively diverse, with over 40% of titles being published outside North America. SA was queried for abstracts fitting in the most granular categories of the DCL (cf. Vancauwenbergh and Poelmans, 2019a).

The main advantage of the service provided by SA, is the acknowledged Thesaurus of Sociological Indexing Terms (Booth, 1996; Blaemers, 2006). This thesaurus allows for structured and specific queries. Additionally, it is possible to determine publication dates, publication types, peer-review status and publication language. For this experiment, we have limited ourselves to peer-reviewed English language journal articles published in the period 2000-2018. To control for adequacy, the queries were manually performed between December 6th (2018) and December 24th (2018).

We sorted the results by relevance, visually inspected the top 100 results to further ensure the accuracy of the results, and downloaded the metadata for the first 1,000 articles. If the query returned less than 1,000 results, we downloaded all articles. The metadata which we retrieved from SA include information on: publication title, author names, journal title, journal ISSN, full abstract, unique identifier assigned by SA, etc. The downloaded metadata were then labelled with the discipline classification codes. In accordance with the number of categories, 77 queries were conducted. This resulted in a dataset with 66,251 entries.

4.2.2 Data cleaning and processing

While exploring the data in preparation of processing, some anomalies were discovered. To account for these, abstracts with more than 600 or less than 60 words, as well as documents with a publication date before 2000, were omitted. After cleaning and labelling the data, we retained 48,961 labelled abstracts. On average, each category contains 635.86 labelled abstracts (min. = 143, max. = 924).

As ML algorithms typically demand for a numeric matrix as input, in our case, the 'features' or columns of this matrix are representations of the words in the abstracts of the publication. These features were first tokenized making use of natural language processing techniques. NLTK, a natural language toolkit implemented in Python, was used. First, we removed English language stop words and performed snowball stemming. Subsequently, the scores contained in the cells of the matrix were vectorised with TF-IDF. Different parameter settings for the ML models were tested making use of Hyperopt (Bergstra, Komer, Eliasmith, Yamins, & Cox, 2015).

4.2.3 Machine learning (ML)

ML approaches come in diverse forms. A distinction can be made between supervised and unsupervised approaches. In this experiment we make use of supervised learning, whereby the algorithm learns from a training data set of abstracts correctly labeled as belonging to a sub-discipline. In this experiment, we evaluate four different supervised ML algorithms, namely: Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM), Random Forest Classifier (RFC), and Gradient Boosting (GB).

The four algorithms were chosen based on their popularity and proven success when implemented in similar scenarios. The implementation of the first three algorithms was carried out with the Scikit-learn package (version 0.20.1) in Python. Scikit-learn harnesses a broad set of ML algorithms (Pedregosa et al., 2011). Given its consistency and relative ease of use, it enables comparison of different applications. For gradient boosting we used the LightGBM implementation (Ke et al., 2017). Accuracy of the algorithms is measured when the models created by the algorithms are fitted to 'unseen' data (i.e. the test data). This score is calculated by dividing the number of correctly categorized documents by the total number of documents.

4.3 Preliminary results

Both MNB and SVM perform relatively poor when compared to the ensemble classifiers (i.e. RFC and LightGBM). While this is not a surprise in itself, the degree to which these latter two classifiers outperform the former is worth noticing. RFC is almost twice as accurate as MNB. SVM outperforms MNB, but is considerably less accurate than RFC and LightGBM. Figure 2 presents a heat map of the results obtained by the LightGBM algorithm. The rows depict the true labels, and the columns depicts the labels predicted by the algorithm. The diagonal in this diagram represents the number of cases that were correctly assigned.

Algorithm	*Achieved accuracy on test set*
Multinomial Naïve Bayes (MNB)	0,49
Support Vector Machine (SVM)	0,68
Random Forest Classifier (RFC)	0,71
Gradient Boosting (LightGBM)	0,82

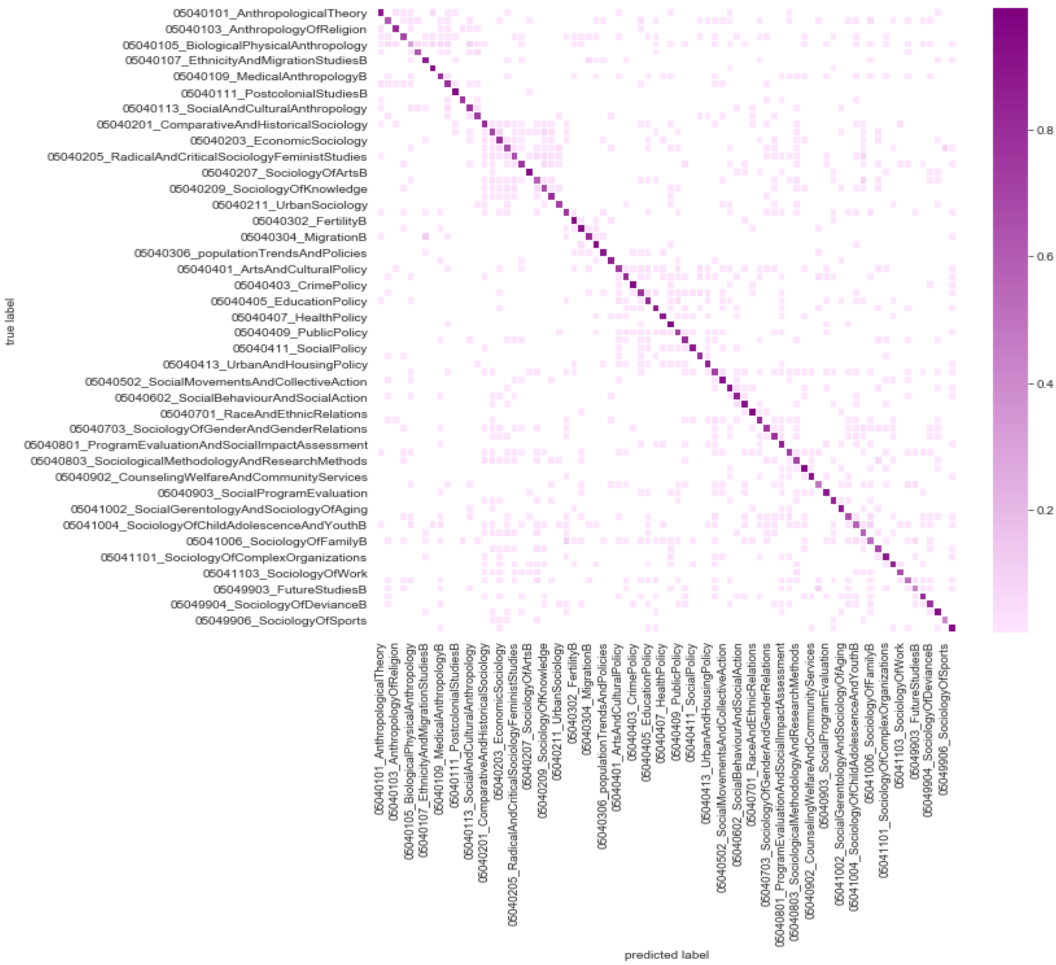
Table 1. Accuracy results for each classifier.

4.4 Conclusion

Our results show that LightGBM can be a fruitful approach to overcome difficulties with regard to a granular classification of scientific articles in SSH. The algorithm was able to correctly classify over 80% of the abstracts collected from SA. Our visualizations (cf. figure 2) show that if mistakes were made, the publications were mostly assigned to neighboring specialties.

A distinction can be made between single- and multi-label TC. The former involves classifying each abstract into one discipline, while in the latter case abstracts may be classified as belonging to multiple disciplines. Whereas the latter would be a more natural way of approaching scientific abstracts (i.e. more often than not, in one single document, there exists a significant overlapping of topics or disciplinary perspectives thereon), in this experiment we focused on single label classification. An exploration of multi-label approaches would be appropriate.

Figure 2. Heatmap of LightGBM results: predicted labels test set (20 % of data) in the columns, true labels in the rows.



5 Scaling up: Applying the system to multiple domains and multi-label documents

Full reference: Eykens, J., Guns, R., and Engels, T. C. E. (2021). Fine-grained classification of social science journal articles using textual data: a comparison of supervised machine learning approaches. *Quantitative Science Studies, 2*(1), 89-110. doi: https://doi.org/10.1162/qss_a_00106

Disciplines have since long been considered as fundamental units of division within the sciences (Stichweh, 2003). These units are knowledge production and communication systems, and can as such serve important classificatory functions (Hammarfelt, 2018; Stichweh, 1992, 2003; Sugimoto & Weingart, 2015; van den Besselaar & Heimeriks, 2006). The subjects of interest for scientometricians, i.e. scientific documents, are classified according to disciplines in order to facilitate research into knowledge production and dissemination. Over the past few decades however, we have faced a continuous growth of the number of new disciplines and specialties (i.e. internal differentiation), resulting in increasing dynamism and 'intensification of the interactions between […] disciplines' (Stichweh, 2003, p. 85).

General classification systems like the Web of Science (WoS) Subject Categories (SC) or the OECD's Fields of Science are too broad to adequately capture the more complex, fine-grained cognitive reality. Several concerns have been raised in this regard—here we mention two central ones. First, Glänzel, Schubert, and Czerwon (1999) point out that the SC approach on the journal level works well for classifying publications in highly specialized journals, but that it is problematic for those appearing in multidisciplinary or general journals. Second, research like Waltman and van Eck (2012) large-scale clustering study, grouping publications based on their citation relations, indicated the feasibility of more fine-grained classification schemes. The authors cluster documents on three different levels, the most detailed of which can be conceived of as 'small subfields' and consists of 21,412 clusters. While most bibliometric studies still make use of more general classification schemes for publications, these are limited in scope, only indicating broad scientific fields or general disciplines. Empirical studies like the one conducted by Waltman and van Eck (2012) as well as theoretical arguments raised by sociologists of science amplify the need for fine-grained classification schemes. More recently, Sjögårde and Ahlgren (2018, 2019) have shown that fine-grained specialized communities can be determined based on citation relations, and these communities in their turn might possibly exhibit specific citation and publication practices.

In Flanders, the Dutch speaking region of Belgium, the Flemish Research Discipline Standard ("Vlaamse Onderzoeksdiscipline Standaard" or VODS) has been introduced to facilitate a detailed classification of research, including research output (Vancauwenbergh & Poelmans, 2019a, 2019b). The VODS builds upon the OECD Fields of Science (2007), adding two more fine-grained levels. While the introduction of the VODS will open up new possibilities for understanding knowledge production and dissemination on a more detailed level, it also poses important challenges, the classification of publications in the social sciences being one of them.

Current bibliometric approaches to classification of publications are not entirely fit for the social sciences. This mainly has to do with lack of coverage in major citation databases (Ossenblok, Engels, & Sivertsen, 2012) and differences in publication and citation practices within the fields (Kulczycki et al., 2018; Nederhof, 2006). One possible way to address these concerns is including non-source items in citation-based bibliometric maps (Boyack & Klavans, 2014). An alternative solution is making use of text-based methods.

5.1 Using textual data and machine learning to cluster or classify (social science) publications

Compared to classification approaches making use of reference/citation data (or other metadata), the usage of purely textual information (i.e. titles, abstracts, full-texts, etc.) has thus far received less attention. Nevertheless, the theoretical relevance of an article's textual content for this task has already been emphasized since the seminal work by Rip and Courtial (1984). Michel Callon and colleagues have further developed this long tradition of co-word analysis research which aims to map and describe scientific interaction and the formation of specialist communities (Callon, Courtial, & Laville, 1991; Callon et al., 1983). More recently there has been a resurgence in interest for textual data, mainly due to increased computing resources and availability of potential data sources.

Machine learning methods currently spearhead a lot of research which is based on textual data. We can distinguish between supervised and unsupervised approaches. In unsupervised learning, no predefined classes or categories are available to learn from. Supervised learning, on the other hand, starts from a set of predefined categories, each of which has a number of instances or records assigned to it. An algorithm is then trained on these labelled instances, from which it tries to deduce the common characteristics of instances in each category, in order to predict to which category a new, unseen instance might belong. The present article uses such a supervised approach.

In scientometric studies, unsupervised clustering of documents is common. Hybrid approaches to document clustering in which citation information and textual data are used have shown that adding textual information can ameliorate the outcomes of document clustering (see for example: Janssens, Zhang, De Moor, & Glänzel, 2009; Yau, 2014). Unsupervised clustering of documents based only on textual similarity (Boyack et al. 2011) has gained traction in the bibliometric community as well. Arguably, supervised machine learning (ML) has been less popular, presumably because in most scientometric clustering studies a granular ground truth classification on the article level is lacking.

An exploration of supervised ML algorithms combined with basic NLP techniques has been described by Read (2010). Read (2010) used supervised learning to classify documents in, among others, the Ohsumed dataset, part of MedLINE. The author reports F1 scores for different multi-label classification techniques, ranging from 0.1 up to 0.43. Classifier Chains are proposed by Read (2010) as a possible solution to the task of multi-label, multi-class classification problems. The latter are tasks in which a document can be assigned to multiple categories at the same time. This kind of learning task is considerably more challenging than the single label classification problem.

Recent supervised ML algorithms with neural networks and word embeddings or BERT (Bidirectional Encoder Representations from Transformers) models respectively, have also been used to vectorize and classify scientific documents. While these recent studies do not deal with multi-label, multi-class classification, they are relevant in that they apply these relatively new NLP techniques to vectorize scientific publications. Kandimalla, Rohatgi, Wu, and Lee Giles (2020) report on a large scale classification study in which they categorize papers according to WoS Subject categories by making use of neural networks and word embedding models. The authors show that such classification systems work well, achieving an average F-score of 0.76. For the individual subject categories the scores range from 0.5 to 0.95. In this study, however, the subcategories with too few records are merged or omitted from the analysis, as they "decrease the performance of the model". Documents which are labeled with more than one category are also dropped. The authors conclude that their experiment shows that the supervised learning approach scales better than citation clustering-based methods. Dunham, Melot, and Murdick (2020) train SciBERT classifiers on arXiv metadata and subject labels. This model is then used to identify AI relevant publications in Web of Science, Digital Science Dimensions and Microsoft Academic. The authors report F1 scores ranging from 0.59 to 0.86 for the 4 categories within the field of AI.

Annif, an automated subject indexing tool currently being tested and implemented at the National Library of Finland, is also comparable to our approach (Suominen, 2019). Annif annotates terms from different subject vocabularies and thesauri to documents based on textual information, like for example abstracts and/or titles. The machine learning module consists of an ensemble of classification algorithms. Annif annotates documents on a granular level, as the tested module was able to assign (up to 5) indexing terms to documents. The module was evaluated on four corpora, including both academic and non-academic texts, yielding F1 scores ranging from 0.14 to 0.46.

The present paper is an extension of work presented at ISSI 2019, where we applied supervised machine learning to classify sociology publications into subdisciplinary categories (Eykens, Guns, & Engels, 2019), reaching 81% accuracy. Note, though, that that paper only worked with publications assigned to one specialty. In this article, we study the use of textual data to classify publications from three social science disciplines into one or more subdisciplines. Much like Read (2010) and Kandimalla et al. (2020), we thus primarily aim to exploit textual characteristics of (social science) documents in order to categorize them into predefined disciplinary categories. As we will describe in more detail further on, we aim to categorize these social science abstracts into granular subcategories. Multiple categories can be assigned to one document at the same time. The novelty of this study resides in the fact that we have used a procedure to validate the data collected for our machine learning experiment, and that multiple granular subdisciplinary categories can be assigned to one single document.

Outline

Under section 5.2 we describe the classification scheme used in detail. Section 5.3 describes the data sources used (i.e. Sociological Abstracts, ERIC, EconLit), and the collection and processing procedure. We have developed a structured way of collecting and validating textual data based on well-established disciplinary thesauri in tandem with a validation round by experts from the respective fields. This validation procedure will be discussed under section 5.3.3. Next, section 0 further details the supervised machine learning algorithms and feature extraction techniques that we compare. Section 5.6 describes the results of the comparison, where we evaluate performance on two dimensions; (i) of the individual labels, and (ii) of the instances. Finally we discuss our machine learning set-up and contrast our approach to existing automatic classification techniques. We conclude with some reflections, pathways for future research, and briefly discuss practical applications.

5.2 The Flemish Research Discipline Standard (VODS)

We make use of the Flemish Research Discipline Standard ("Vlaamse Onderzoeksdiscipline Standaard", abbreviated VODS, in Dutch), which is available at https:// researchportal.be/en/disciplines and has been described in the literature (Vancauwenbergh & Poelmans, 2019a, 2019b). The VODS has been introduced in the Flemish Research Information Space (FRIS, see https://researchportal.be/en), an aggregation platform of publicly funded research in Flanders, in 2019. In the future, all scientific output produced by scholars in Flanders might be classified according to the VODS. The VODS is structured as a hierarchical tree with four levels. To allow for international comparison, the first level overlaps with the 7 broad fields of science present at the highest level of the OECD Fields of Science (OECD, 2007) coding scheme (hereafter referred to as OECD FOS). For the case of sociology, for example, on the top level of the OECD FOS we find category 5 'social sciences' and subcategory '5.4 sociology and anthropology' (Figure 1). This category is present in the VODS as well.

Hierarchical structure of OECD FOS (ex. 5.4)	**Hierarchical structure of VODS (ex. 050405)**
5. Social Sciences	05 Social Sciences
5.4 Sociology	0504 Sociology and Anthropology
	050401 Anthropology (14 subcategories)
	050402 Applied Sociology (11 subcategories)
	…
	050405 Social Change
	05040501 Social change
	05040502 Social movements and collective action
	05040599 Social change not elsewhere classified
	…

Table 2. Excerpt of tree structure: OECD FOS (2007) coding scheme and VODS (2019). The VODS classification scheme can be accessed at https://researchportal.be/en/disciplines.

The VODS adds two more granular layers representing further subdivisions of the second layer of the OECD FOS. The third layer of the VODS might be interpreted as containing sub-disciplinary categories, while items on the fourth level can be considered research specialties. To construct and define this scheme, experts from the corresponding fields were consulted by the creators of the VODS. In total, on the most granular level the VODS contains 2,493 codes. For further technical details on this classification scheme, we refer interested readers to Vancauwenbergh and Poelmans (2019a, 2019b).

Our objective is to automatically classify articles (based on abstracts and titles) into categories on level 3 of the coding scheme (e.g. 050402 Applied sociology and/or 050405 Social Change and/or …) for three fields within the social sciences, namely (0502) economics & business (10 classes on the third level), (0503) pedagogical & educational sciences (9 classes on the third level), and (0504) sociology & anthropology (12 classes on the third level). On level 3, we have 31 sub-disciplinary categories for the three disciplines together. Section 5.3.3 will further detail the reasons why our approach operates on level 3 rather than level 4. In the following part we introduce the data sources used to collect the titles and abstracts for the three disciplines.

5.3 Data sources: Sociological Abstracts, ERIC and EconLit

The data used for our study were downloaded from ProQuest (https://search.proquest.com). ProQuest provides good journal coverage of the social science literature compared to, for example, Scopus or WoS (Norris & Oppenheim, 2007). ProQuest offers access to a range of existing abstracting services and disciplinary databases. For the purpose of our analyses we have used Sociological Abstracts to download bibliographic records from sociology & anthropology, EconLit for records from business & economics, and ERIC for records from the pedagogical & educational sciences.

5.3.1 Combinations of indexing terms as proxies for subject specialties

A clear advantage of all three databases is that they make use of controlled vocabularies (or thesauri) for the records which are indexed. The Thesaurus of Sociological Indexing Terms is a well-developed and highly regarded indexing system used by Sociological Abstracts' service. Within EconLit, the Journal of Economic Literature (JEL) classification, also known as the American Economic Association Classification System, is used. Within ERIC, the Thesaurus of ERIC Descriptors is used. In addition, ProQuest's search engine allows to filter on publication types and publication years. We selected all journal articles published between 2000 and 2018. These controlled vocabularies allow us to query ProQuest's command line search page for abstracts on a very fine levels of granularity. Figure 3 shows an example of the query we used for category 'law & economics', within business & economics. The full set of queries for all categories is available online (Eykens & Guns, 2020).

Figure 3. Example of command line query designed for Business & Economics, category 05020109.

```
MAINSUBJECT.EXACT("Personal Bankruptcy Law (K35)")
  OR MAINSUBJECT.EXACT("Regulation and Business Law: General (K20)")
  OR MAINSUBJECT.EXACT("Regulation and Business Law: Other (K29)")
  OR MAINSUBJECT.EXACT("Tax Law (K34)")
  OR MAINSUBJECT.EXACT("Regulated Industries and Administrative Law (K23)")
  OR MAINSUBJECT.EXACT("Regulation and Business Law (K2)")
  OR MAINSUBJECT.EXACT("Business and Securities Law (K22)")
  OR MAINSUBJECT.EXACT("Contract Law (K12)")
  OR MAINSUBJECT.EXACT("Tort Law and Product Liability; Forensic Economics (K13)")
  OR MAINSUBJECT.EXACT("Law and Economics: General (K)")
```

To design the queries, we manually coupled the indexing terms of these different datasets to the fourth level categories in the VODS and downloaded the abstracts found for each query. The first author went through the list of VODS categories and per category collected all relevant indexing terms from the thesaurus at hand.

The VODS provides semantic definitions of each category, which were formulated together with field experts. We used this information to manually retrieve the relevant indexing terms. In many cases, this was straightforward, because there was a perfect overlap with the indexing terms (for EconLit, this was the case for nearly all categories). In other cases some additional indexing terms were found to be relevant (see for example Figure 3).

The indexing terms were then used to query ProQuest. The records retrieved for each of the 214 level 4 categories were subsequently downloaded (with an upper limit set to 1,000 records per VODS level four category) and saved in a separate folder which was labeled with the corresponding VODS category. The data collection was carried out between December 2018 and February 2019. After collecting and processing, the merged sets (all files for the three fields together) resulted in a raw set consisting of 148,341 records (see Table 3).

5.3.2 Data cleaning and processing

To clean the raw dataset, we followed a protocol consisting of four steps. First (1) we removed all records that were missing an abstract or title. Although we limited our search to records published between 2000 and 2018 (2), there were still some in our dataset which were published before 2000 or after 2018. These were omitted as well. Lower and upper boundaries were set for the word count of the abstracts (3), minimum 50 and maximum 1,000 respectively. These limits were found to adequately weed out cases where the abstract field replicated either the title or the entire full-text.

Whereas we expected only journal articles resulting from our queries, other publication types were present as well. The reason for this might have to do with the fact that all three datasets have been designed by different organizations, which results in a diverse range of variable names to describe the different publication types used within the datasets. (4) For each dataset, we compiled a list of unique variable names present in the collected records and filtered out those describing publication types that we did not want to take into consideration (e.g. book reviews, interviews, editorial material, instructional material, etc.). A list of the remaining, relevant publication types was used to restrict our dataset to research articles published in journals.

Table 3 provides an overview of the number of records in each set before and after cleaning. For Sociological Abstracts and EconLit, our initial collection of records was reduced by a little over 20%. For ERIC, the total number of records was reduced by almost 40%. The large intergroup difference observed is mainly due to a large number of records classified as 'instructional material' in ERIC. The large intragroup difference is due to the smaller number of subcategories present in the VODS. For business & economics we queried 84 categories, for pedagogical & educational sciences 53 VODS categories, and for sociology & anthropology 77 categories.

As discussed above, we have designed queries for each level 4 category in the VODS and collected records from the respective databases. Some records appeared multiple times – that is, some records were retrieved with different queries. After deduplication and relabeling, the dataset contains 113,909 multi-labeled abstracts, with an average of 1.1 labels per abstract (min. = 1, max. = 6, SD = 0.36).

5.3.3 Expert validation: Inter-indexer consistency and F1 scores

In order to validate the reliance on controlled vocabularies described above, a domain expert from each of the three disciplines was contacted. The three experts were given a random sample of 45 abstracts and titles, which they were asked to classify according to the VODS level 4 categories corresponding to their field of expertise (i.e., sociology & anthropology, business & economics, and pedagogical & educational sciences). Each expert was presented a set of abstracts and titles from their own discipline. The expert working in the field of business & economics, for example, was given abstracts and titles originating from EconLit (business & economics) only. No limitations were set on the number of categories the indexers were allowed to assign.

Next, the classification by the experts on the one hand and the classification based on the controlled vocabularies of each database on the other were compared. In order to do so, the inter-indexer consistency (IIC) was calculated for each record in the sample. For every sample, we calculated the average IIC using the method described by Rollins (Leininger, 2000). First, a percentage of consistency between two indexers (here: the expert and the controlled vocabulary) is calculated for each document d:

$$IIC_d = \frac{2A}{B + C} \#(1)$$

Here, A denotes the number of categories on which both indexers agree, *B* is the number of categories assigned by indexer 1 (expert) and *C* the number of categories assigned by indexer 2 (controlled vocabulary). The IIC at document level is the Dice coefficient of the two sets of categories assigned by the indexers. The average IIC for the whole sample is calculated by dividing the sum of the IICs for all individual documents by the total number of documents N (in our case, equal to 45). In addition, we calculated F1 scores for each disciplinary sample. We have calculated these scores for level 3 and 4 of the VODS.

Table 4 displays the results of the IIC and F1 calculations**.** The F-scores are also included for the assessment of the performance of the machine learning models. On level three of the VODS, the IIC varies between 45.2% for the sample from Sociological Abstracts and 62.2% for EconLit. On level four, the IIC scores are considerably lower, with a minimum of 23.7% for EconLit and a maximum of 39.7% for ERIC. Previous research into IIC in the case of the PsycINFO database shows similar results with those obtained for level 3 of the VODS. Leininger (2000) evaluates IIC for a similar classification scheme, based on research areas within psychology. Using Rollins' method, he finds an average IIC of 45% (Rollin, 1981, as cited in Leininger, 2000, p. 6). Sievert and Andrews (1991) study the IIC for a subset from Information Science Abstracts. The authors report average consistency scores of about 50%. Funk (1983) study the IIC for MEDLINE. They report a consistency score of 61.1% for the MeSH terms assigned to documents. While our scenario is somewhat different, i.e. the first author 're-classified' publications according to the indexing terms and experts were consulted to validate this reclassification, it seems that these low scores rather indicate the difficulty of the problem at hand. Therefore, we conclude that the level 3 classification is sufficiently robust to be used for our supervised machine learning experiment and hence we limit ourselves to the classification of journal articles at this level. For matters of interpretation of the differences in scores, on level 3 we have 31 subdisciplinary categories in total, compared to 214 research specialties on level 4.

VODS category	Indexing service consulted	Initial number of records	Number of records after cleaning
0502 Business & economics	EconLit	63,407	50,577
0503 Pedagogical & educational sciences	ERIC	23,521	14,527
0504 Sociology & anthropology	Sociological Abstracts	61,413	48,805
Total		**148,341**	**113,909**

Table 3. Number of records collected from each database: before and after cleaning.

Sample from	IIC level 3	F1 level 3	IIC level 4	F1 level 4
Pedagogical & educational sciences – ERIC	52.9%	0.59	39.7%	0.42
Business & economics - EconLit	62.2%	0.57	23.7%	0.51
Sociology & anthropology - Sociological Abstracts	45.2%	0.67	26.7%	0.48

Table 4. Rollins (1981) inter-indexer consistency (IIC) and weighted F1 scores for the three datasets on two classification levels.

5.4 Methods

Since 30 years, the dominant paradigm of text classification (TC) consists of machine learning (ML) approaches. ML algorithms are deployed such that "a general inductive process automatically builds an automatic text classifier by learning, from a set of preclassified documents, the characteristics of the categories [or labels] of interest" (Sebastiani, 2002, p. 2). ML approaches have already been applied to classify (abstracts or full-texts of) journal articles. Langlois, Nie, Thomas, Hong, and Pluye (2018) classify papers into two broad domains, namely empirical and non-empirical. Our approach is different from such studies as the level of granularity of categories into which we classify texts is far greater. Consequently, articles from two different level 3 subdisciplinary categories are overall much more similar than what is encountered in most other classification tasks.

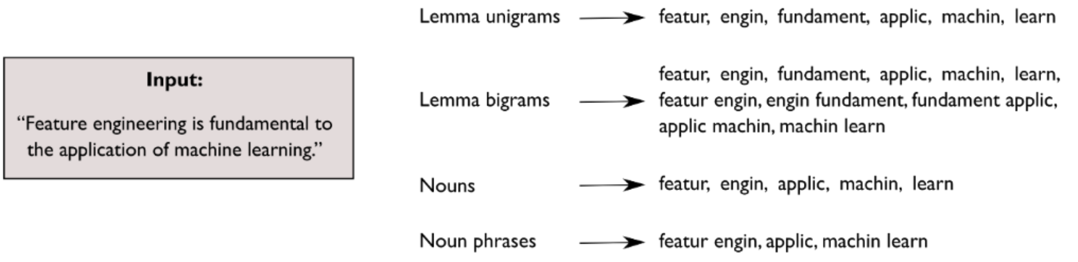
The classification problem discussed in this paper belongs to the domain of multi-class multi-label classification. Multi-class classification refers to assigning one out of more than two classes to an instance. Multi-class multi-label classification is an extension of this problem where we assign one or more out of multiple classes to an instance (Read, Pfahringer, Holmes, & Frank, 2009). Some abstracts were thus assigned to multiple classes (up to a maximum of 6).

A popular strategy is to transform the multi-label problem into different single-label classification tasks. This can be done making use of binary relevance. As a baseline classifier, we make use of Multinomial Naïve Bayes (MNB). We optimize this classifier in order to explore the best feature engineering techniques as described below. Next, we compare the results obtained with MNB to those obtained by a Gradient Boosting model. After discussing the feature engineering steps in the following part, we will present a short description of the algorithms and the metrics that were used to evaluate performance on different aspects.

5.4.1 Feature engineering

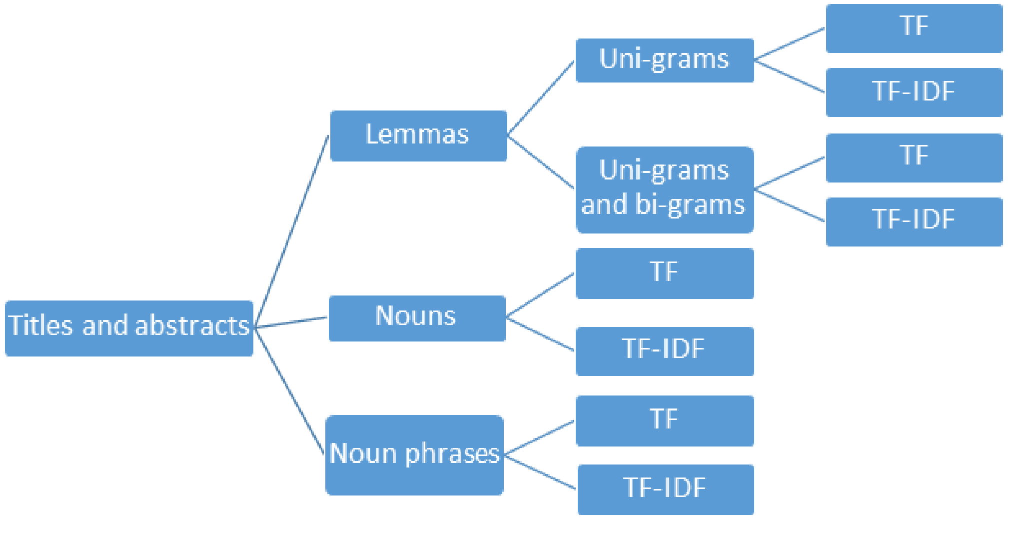
Feature engineering for multi-label TC is done in the same way as for single-label TC. The 'features' or columns of the matrix are representations of words in the abstracts and titles of the publications. The Bag of Words approach (BoW) is a traditional, popular, and simple yet powerful way of vectorizing documents for TC. The BoW approach consists of slicing a text into words or phrases (without taking word order into account). We have built customized tokenizer functions in Python to extract four different textual features: lemma unigrams, lemma bigrams (combined with unigrams), nouns, and noun phrases (cf. Figure 4). Although previous research has shown that for the BoW approach more advanced document representations like nouns and noun phrases are 'not adequate to improve TC accuracy', we wanted to explore this for our specific use-case (Moschitti & Basili, 2004).

Figure 4. Four feature extraction methods: lemma unigrams, lemma bigrams, nouns, and noun phrases.



We made use of the natural language processing packages NLTK (Loper & Bird, 2002) and SpaCy (Honnibal & Montani, 2018) to parse the texts, perform part of speech tagging and stemming. For stemming, we made use of NLTK's implementation of the snowball stemming algorithm. Scikit-learn's count vectorizer and TF-IDF (term frequency-inverse document frequency) transformer were used to process the outcomes of the different feature extraction methods (Pedregosa et al., 2011). With each tokenizer, we tested the performance of both (normalized) TF and TF-IDF. This resulted in 8 different feature spaces (see Figure 4).

Figure 5. Overview of feature transformation steps.



Feature sparseness is a common problem in TC. Transformation methods which make use of bigrams can easily bring about feature matrices with hundreds of thousands or even millions of columns, leading to a very high dimensionality. To reduce the dimensionality, we make use of a feature selection method based on randomized decision trees. After extracting textual features we fit a shallow extra trees classifier (maximum depth of 10) to the data in order to select the most relevant ones.

5.4.2 Classification algorithms

Multinomial Naïve Bayes (MNB)
MNB is one of the most popular TC algorithms used by the ML community. It is a fast, scalable (i.e. iterates very fast over large datasets) and successful approach for many TC problems. Over the years, it has become a popular baseline method, 'the punching bag of classifiers' (Lewis, 1998, p. 2). MNB makes use of Bayes's theorem to construct histograms based on the feature vectors – in our case counts or probabilities of the textual features present in a document – for every single instance. The classifier associates these histograms with the labels, and estimates likelihoods of a label and a distribution of feature counts occurring together.

If, however, a feature-class combination has zero counts, the probability will be set to zero. This mitigates the necessary information of the other probabilities by multiplying them by zero. For the algorithm to be able to deal with such problems a smoothing parameter is used. Another way of dealing with this problem is transforming the feature space into a TF-IDF normalized matrix (Rennie, Shih, Teevan, & Karger, 2003).

Gradient Boosting Decision Trees (LightGBM)
Gradient Boosting Decision Trees (GBDTs) are, as the name indicates, tree-based learning algorithms. These algorithms build ensemble models, or groups of decision trees aimed at reducing residual errors for a split point in a decision tree. Boosting is a specific ensemble technique, which sequentially builds the models on random subsets of the features and instances. When an instance is misclassified, its weight is increased and the next model tries to correct for this error.

In practice, this algorithm can be very time-consuming. Ke et al. (2017) have come up with a solution to this problem by optimizing the randomness of the feature and instance selection step. They combine Gradient-based One-Side Sampling (GOSS) with Exclusive Feature Bundling (EFB) in order to speed up the training process. The GOSS procedure pays more attention to instances with larger gradients (i.e. having more impact on the classification error of a model) 'and randomly drop those instances with small gradients' (Ke et al., 2017, p. 2). This approach is implemented in the LightGBM software package (https://lightgbm.readthedocs.io).

The EFB implementation exploits feature sparseness, which is a very common problem in text classification. It bundles sparse features together into a single feature, efficiently reducing the dimensionality. In a previous study, we have found (the LightGBM implementation of) gradient boosting to be the best performing algorithm to classify publications in sociology & anthropology, achieving accuracy scores well over 80% (Eykens et al., 2019). Different from this previous study, in this paper we assess classifier performance for a vastly more complex multi-label setting.

Decision tree-based models, however, come at a cost. They require tuning a wide range of parameter settings. For LightGBM, one can set well over 100 parameters[13]. For our purposes, we have chosen to optimize for 11 core parameters:

- the number of trees that will be built;
- the maximum depth of the trees: to limit tree growth;
- the number of leaves of the decision trees: last splits made in the model when reaching the optimal number of splits for a given loss function, or when reaching the predefined maximum depth;
- the learning rate: sets the weight of the outcomes of each tree for the final output;
- maximum in bin: handles the maximum number of bins in which the feature values will be grouped;

[13] For a complete overview of the parameters used in LightGBM, see https://lightgbm.readthedocs.io/en/latest/Parameters.html

- regularization alpha (L1): limits the impact of the leaves encouraging sparsity (i.e. weights to zero);
- regularization lambda (L2): limits the impact of the leaves by encouraging smaller weights;
- minimum child weight: the minimum sum of instance weight which is needed in a leaf (child);
- bagging fraction: the fraction of the dataset used for each iteration;
- bagging frequency: the number of trees training per random subsample of the dataset;
- minimum data in leaf: the minimum number of samples which should be captured in a leaf.

We will describe how we optimized the parameters in section 5.4.3.

Multi-label classification: Classifier Chains (CC)
Two main approaches to multi-label classification exist: problem transformation and algorithm adaptation. The most popular and computationally least expensive approach is problem transformation, where a multi-label classification problem is transformed into N single label classification problems. An example of problem transformation is turning the multi-label task into N-labels binary classification problems, wherein each binary classification problem is treated by a separate classifier. This is also known as the Binary Relevance method and has proven success in the domain of multi-label TC (Zhang, Li, Liu, & Geng, 2018).

As each label is treated separately, however, the algorithm effectively ignores label dependence. Read et al. (2009) have suggested an improvement of the Binary Relevance method by 'chaining' the results of each classifier to the input space so that the next training round takes the results of previous classifiers into account. As different disciplinary categories might be closer to each other in terms of concepts and topics studied, we do not expect labels to be completely independent of each other. Hence, we opt for the Classifier Chains (CC) approach. It should be noted that other approaches exist, but these come at a cost of computational complexity as well as intuitive understanding of the models.

5.4.3 Cross-validation

After vectorization, dimensionality reduction and problem transformation with a binary relevance based Classifier Chains (CC) algorithm, a hold-out set (25% of the complete dataset) was sliced from the initial dataset using an iterative stratification technique as proposed by Sechidis, Tsoumakas, and Vlahavas (2011). This stratification method handles class imbalance for multi-label learning problems in such a way that the distribution of instances over classes in the validation set is kept as close to the actual distribution as possible.

Figure 6. Visualization of training set - validation set folds and test data. Lighter grey are training samples, and darker grey validation samples.

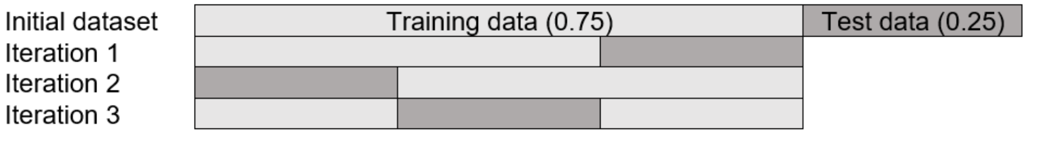


Figure 6 visualizes the cross-validation procedure. The test data (0.25 of total set) will be used for the final evaluation of our models. For each iteration, a different subset of the remaining 75% of the data (training data) are used to evaluate different parameter settings for the feature engineering options presented above. We make use of randomized parameter grid search and three-fold cross validation to evaluate different parameter settings on parts or 'folds' of the training data. This means we run three new random experiments, each of which again divides the training data into two different parts, using 66.66% of the training data to train a model with a random parameter setting, and evaluating that setting on unseen data (the darker grey area represented above). We make use of three different slices of training and test data to make sure our findings are robust.

5.5 Evaluation metrics

Evaluating the performance of multi-label classification is not as straightforward as is the case for single-label classification. Single-label classifiers' predictive performance can be evaluated using the accuracy measure, i.e. the fraction of correctly classified instances over the total number of instances. The Accuracy Acc is calculated as follows:

$$Acc = \frac{1}{N}\sum_{i=1}^{N} I\left(Y_i = \hat{Y}_i\right) \#(2)$$

I here is the indicator function. Y_i is the set of true labels (sub-disciplinary categories, in our case) for document i, and $\hat{Y}_i$ is the set of predicted labels for document i. For multi-label classification assessing such a score based on the full set of labels per instance would be too harsh, "since even a single false positive or false negative label makes the example incorrect" (Read, 2010). Using multiple metrics to capture different dimensions of the multi-label prediction is advised (Read, 2010; Zhang & Zhou, 2014). Two main dimensions can be assessed: the individual labels and the entire training or testing label sets per instance (Zhang & Zhou, 2014, p. 1822). For full label set evaluation, we calculate accuracy, and for label-based evaluation, we calculate precision and recall.

Per label ℓ from the set of labels L, we can determine the set test_ℓ of documents to which this label has been assigned and the set pred_ℓ of documents for which the classifier predicts this label. Weighted average precision P is determined as follows:

$$P = \frac{1}{\sum_{\ell \in L}|\mathrm{test}_\ell|}\sum_{\ell \in L}|\mathrm{test}_\ell|\frac{|\mathrm{test}_\ell \cap \mathrm{pred}_\ell|}{|\mathrm{pred}_\ell|} \#(3)\#\#$$

where $|\cdot|$ denotes set cardinality. Similarly, weighted average recall R is:

$$R = \frac{1}{\sum_{\ell \in L}|\mathrm{test}_\ell|}\sum_{\ell \in L}|\mathrm{test}_\ell|\frac{|\mathrm{test}_\ell \cap \mathrm{pred}_\ell|}{|\mathrm{test}_\ell|}$$
$$= \frac{\sum_{\ell \in L}|\mathrm{test}_\ell \cap \mathrm{pred}_\ell|}{\sum_{\ell \in L}|\mathrm{test}_\ell|} \#(4)\#$$

The F1 score is the weighted average of precision and recall. Precision and recall are first 'macro-averaged' by calculating the weighted mean of precision and recall for each label, and these are used to calculate the final F1 scores. These measures give an indication of the performance of our algorithm across the three different disciplinary datasets. Precision (3) in a multi-label setting is “the fraction of predicted relevances which are actually relevant” (Read, 2010, p. 41). In addition, Schapire and Singer (2000, as cited in Tsoumakas and Katakis, 2007) propose Hamming loss to take into account the fraction of labels that are predicted incorrectly. Hamming Loss is calculated as follows (see Sorower, 2010):

$$Hamming\ Loss = \frac{1}{N|L|}\sum_{k=1}^{|L|}\sum_{i=1}^{N} y_{i,k} \oplus \hat{y}_{i,k} \ \#(5)$$

Here, $\oplus$ is the exclusive-or operator, $y_{i,k}$ is 1 if document i has label k and 0 otherwise, and similarly, $\hat{y}_{i,k}$ is 1 if document i is predicted to have label k and 0 otherwise. We average these scores over the total number of classes $|L|$ and predictions N. Hamming Loss thus denotes the fraction of incorrectly predicted labels and its optimal value is 0.

5.6 Results

In the first part, we present the best results obtained for Multinomial Naïve Bayes. As detailed in the above, we have vectorised the abstracts and titles making use of three (slightly different) textual characteristics, namely lemmas, nouns and noun phrases. Because of the computational requirements, the machine learning steps were carried out on the High Performance Computing infrastructure of VSC (the Flemish Supercomputer Center) at the University of Antwerp.

5.6.1 Multinomial Naïve Bayes

For the Multinomial Naïve Bayes classifier, we aim to optimize the smoothing parameter alpha. We randomly sample a value from a loguniform distribution, ranging from very small (i.e. 1e-10) up to 1 (i.e. add-one or Laplace smoothing). After finding the optimal value for alpha (0.13883) by fitting the algorithm to the three folds of the training set, we make a prediction for the held-out test set. The results for the best feature representation method are presented in

Table 5. The optimal representation strategy turns out to be lemma bigrams without IDF normalization.

Making use of bigrams for lemmas decreased the Hamming Loss and increased the other scores. We achieved quite similar results with TF-IDF transformed vectors. Interestingly, noun phrases, except for the Hamming Loss evaluation metric, do not yield improved results.

5.6.2 Gradient Boosting (LightGBM)

For the Gradient Boosting algorithm, we randomly sample values for 11 different parameters. To reduce computing time, we have limited the number of random iterations to 100. If we were to perform a full parameter grid search, the number of model fits would be far too high. Keeping in mind that 25 fits take about three hours, this is not desirable.

Compared to the best results achieved with MNB, the Gradient Boosting implementation scores better on almost all evaluation metrics, except for precision (see Table 6). It is interesting to note that MNB scores better for the precision metric in some scenarios. Accuracy scores however, strongly same feature transformation strategy seems to work best for Gradient Boosting. For the lemma bigrams feature extraction score of 0.55.

	TF				TF-IDF			
	Lemma uni-grams	Lemma bi-grams	Nouns	Noun phrases	Lemma uni-grams	Lemma bi-grams	Nouns	Noun phrases
Set accuracy	0.20	**0.24**	0.17	0.14	0.19	0.21	0.17	0.15
Hamming loss	0.04	0.04	0.04	0.03	0.04	0.03	0.04	**0.03**
Precision	0.61	0.61	0.62	**0.66**	0.60	0.64	0.61	0.63
Recall	0.31	**0.37**	0.26	0.19	0.30	0.31	0.25	0.20
F1 score	0.36	**0.42**	0.31	0.27	0.35	0.38	0.31	0.29

Table 5. Results Multinomial Naïve Bayes classification performance for optimal feature space (lemma bigrams, no IDF normalization). Training and evaluation on hold-out set. Best results (per row) in bold.

	TF				TF-IDF			
	Lemma uni-grams	Lemma bi-grams	Nouns	Noun phrases	Lemma uni-grams	Lemma bi-grams	Nouns	Noun phrases
Set accuracy	0.46	**0.46**	0.43	0.33	0.45	0.45	0.43	0.32
Hamming loss	**0.03**	0.03	0.03	0.04	0.03	0.03	0.03	0.04
Precision	**0.66**	0.64	0.60	0.49	0.66	0.63	0.60	0.49
Recall	0.48	**0.50**	0.45	0.36	0.48	0.50	0.45	0.36
F1 score	0.54	**0.55**	0.49	0.40	0.54	0.55	0.49	0.39

Table 6. Scores for Gradient Boosting classification on the validation set, for each feature space. Best results (per row) in bold.

Hamming loss is considerably lower as well, with a fraction of 0.3% of the labels wrongly assigned. 46% of the label combinations predicted by the algorithm were the same as the ones in the test set. It is noteworthy that the differences between TF-IDF and TF feature transformations are insignificant.

Figure 7. Box plot of F1 scores for all 31 subdisciplines for MNB and Gradient Boosting (GB). Preprocessing: lemma bigrams, no IDF. The subdisciplines are grouped per discipline and the vertical line segments indicate the average F1 scores per discipline.

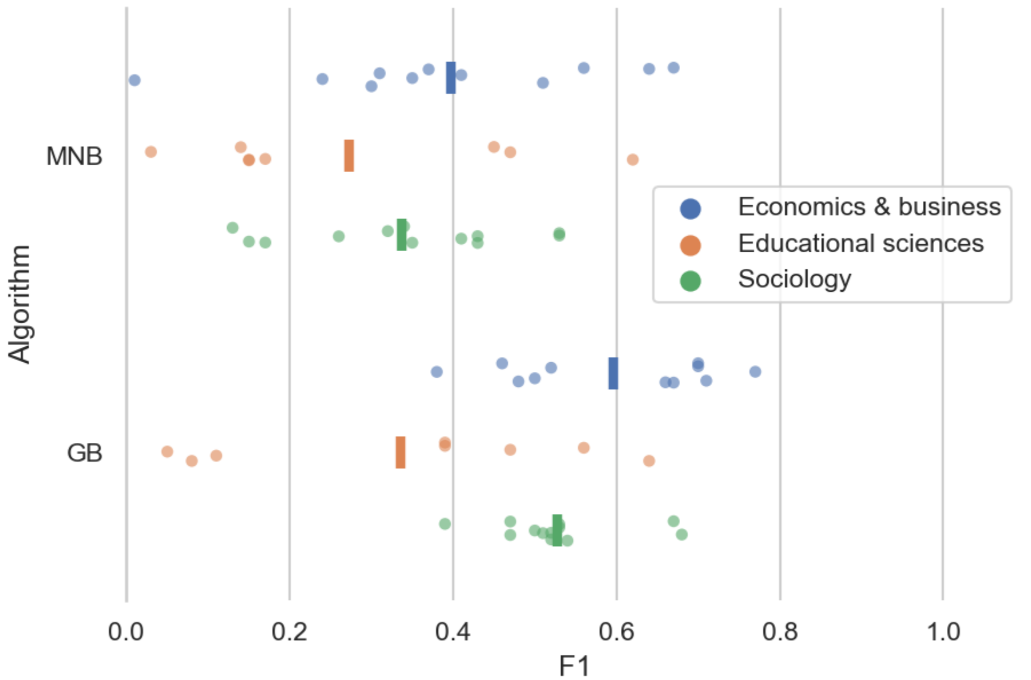
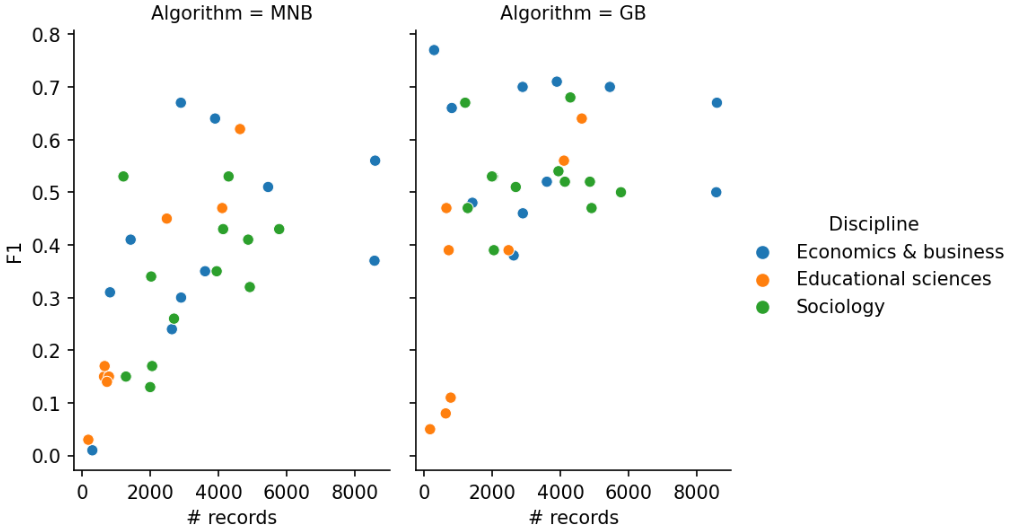


Figure 7 shows how the F1 scores are distributed across all 31 subdisciplines. We observe that the scores for Gradient Boosting are not only higher on average but also less spread out, with the exception of three poorly scoring subdisciplines. These three are all subdisciplines of educational & pedagogical sciences: Informal learning, General pedagogical & educational sciences, and Parenting & family education. Except for these sub-disciplinary categories, overall, no discipline performs clearly better or worse than the others, although the number of training records seems to have some influence: subdisciplines with fewer training records tend to get lower F1 scores (Figure 7). While this relation is somewhat stronger for MNB, the three cases for Gradient Boosting with exceptionally low F1 scores all have few (between 174 and 780) records.

Figure 8. Relation between the number of records and F1 scores for MNB and GB for each of the 31 subdisciplines studied. Preprocessing: lemma bigrams, no IDF.



5.7 Discussion

Classifying research output into disciplinary categories is of fundamental importance for nearly all bibliometric analyses. In the introduction of this paper, we touched upon the issue of differentiation in the sciences, leading to an ever increasing number of research communities and disciplines (Stichweh, 2003). This emergence of new disciplines can be proceeded by, among other things, the formation of new research specialties, the organization of new conferences, the formation of new scientific societies and the foundation of new journals (see Shneider, 2009). As the landscape of disciplines grows more diverse, classification schemes are being updated to better fit this dynamic reality.

The development of such an updated classification scheme is exemplified by the implementation of the VODS in Flanders (see Vancauwenbergh & Poelmans, 2019). Such a diverse and fine-grained classification scheme makes it possible to study interactions between disciplines (i.e. inter- and intra-disciplinary knowledge flows) more closely, and map discrepancies between different classification systems with more detail. Yet, it requires new ways of approaching classification tasks as well, in particular in settings such as the classification of expertise, projects and outputs for which citation data are not available. In this article we take up the specific challenge of a fine-grained classification of social sciences journal articles using the text of their abstracts and titles.

To summarize, our study consists of three elements. First, we constructed a labelled dataset. Since the VODS classification scheme is relatively new, we lack a dataset of classified publications or other documents that can readily be used for machine learning purposes. This led us to manually construct a training dataset consisting of data extracted from EconLit, ERIC and Sociological Abstracts. Each of the 31 VODS subdisciplines of economics & business, pedagogy & educational sciences, and sociology & anthropology was translated to a thesaurus-based query for the respective databases. Second, the query results were validated by human experts. IIC and F1 scores indicate that categories at level 3 (subdisciplines) and 4 (specialties) of the VODS can sometimes be hard to distinguish between. At the same time, the IIC scores for level 3 categories are comparable to those obtained in earlier IIC studies.

Third, the labelled dataset at level 3 was used to train Multinomial Naïve Bayes and Gradient Boosting machine learning models. If we compare Figure 6 to Table 4, the configuration with the best results yields F1 scores slightly below those for the validation by human experts. This indicates that the models might still be improved somewhat, but very high scores are probably unrealistic or indicative of overfitting. Taken together, the results suggest that level 3 of VODS is so fine-grained that some categories are hard to discern in practice and as a result a certain degree of ambiguity becomes unavoidable, at least for the disciplines studied here.

While some of the reported indicators, such as F-scores, are relatively low, we think it is instructive to compare our results to those of the recent studies by Kandimalla et al. (2020) and Dunham et al. (2020). While these authors report better accuracy, it should be highlighted that in this paper we specifically look at the applicability of supervised learning in the context of social sciences. As Kandimalla and colleagues note, this is not an easy task given the large overlap in terminology and the proximity of the categories. Kandimalla et al. (2020) have for that reason dropped or collapsed 120 out of 235 of the subject categories from their dataset. In addition, they drop documents assigned to multiple disciplines. It should be noted that WoS subject categories are less granular than the ones used in our study, i.e. on the level of disciplines instead of subdisciplines. Dunham et al. (2020) report good scores for their model which classifies AI publications into subdisciplinary categories, but their model is restricted to only four categories in AI. Hence it is also less prone to errors. Our system works with 31 subcategories, divided over three social science disciplines. Taking these elements into account, it becomes clear that the lower scores are to a large extent a result of the difficulty of the task at hand.

A matter of concern which can be raised in this regard, is to what extent classification of documents at a level of granularity that is finer than that of disciplines is feasible. Disciplines, and especially subdisciplines and research specialties, are in constant flux. Whereas most publications might belong to the knowledge base of just one discipline, their contents may be of relevance to two or more subdisciplines and research specialties. Theoretical work like actor-network theory in the social sciences, for example, has been of relevance for many disciplines, subdisciplines and research specialties, not only in the social sciences. Interdisciplinary studies, in which an integration of different disciplinary knowledge sources takes place to tackle a research question, may classify under several research specialties, subdisciplines and disciplines. As these examples illustrate, a multi-label approach as applied in this paper is needed in view of the validity of a classification.

This framework requirement needs to be balanced with requirements in terms of accuracy, feasibility and reliability of a classification scheme. As the results of our study show, the classification of social sciences publications into subdisciplines (VODS level 3) on the basis of abstracts and titles is a hard task for both humans and machines; classification into research specialisms (VODS level 4) probably is not all that meaningful anymore (cf. the IIC and F1 scores in Table

4). We argue that classification at the subdiscipline level should be further explored and fine-tuned, as this level of granularity corresponds to actual policy needs and might be improved by smart combinations of human input and ML. For example, a recommender system might be improved through validation by the authors of papers and machine classifications might gain accuracy through the use of larger sets of texts describing expertise, projects and publications classified by humans.

Limitations

Four limitations of this paper should be highlighted. First, we could not compare our results to any benchmark. Although there have been some experiments in which supervised machine learning techniques are used to classify (or study elements of) scientific articles (see for example Langlois et al., 2018; Matwin & Sazonova, 2012), to date no comparable applications or datasets exist (i.e., medium sized, annotated sets of social science publications classified according to fine-grained disciplinary categories) – at least not to our knowledge. The lack of previous work in this line of research makes it hard to benchmark our results for this specific problem setting.

Second, given that the records in our dataset were extracted from EconLit, ERIC or Sociological Abstracts, each record has been assigned to only one (but possibly multiple subdisciplines of the same) discipline of the VODS level 2, i.e., to economics & business, to pedagogy & educational sciences or to sociology & anthropology. Hence interdisciplinary cases are not present in our initial training data. We cannot compare the performance of the models deployed in this study at different levels of granularity, in particular the discipline and the subdiscipline level. However, our results do show that the subdiscipline level is, at least for articles in social sciences and using their abstracts and titles only, the most fine-grained level that makes sense for classification exercises.

Third, we have coupled classification systems with two entirely different functions. On the one hand, we have the indexing systems based on the thesauri. These are systems that are designed for information retrieval purposes and have no limit to the number of indexing terms that can be assigned to a document. In such a system, there is no purpose in trying to fit a document into one to six sub-disciplinary categories. Thus, we have reduced the complexity and granularity of the thesaurus-based classification to a fixed number of disciplinary groups. This 'mismatch' between the two classification systems might lead to relatively low scoring results when a machine learning algorithm is tasked with reproducing this classification.

Fourth, as discussed in the methods section, the queries have been manually constructed by the first author. The indexing terms in the thesauri were coupled to VODS discipline codes based on the semantic definition of each field in the VODS. It can be argued that this is a highly subjective task, as previous research has shown that disagreement between indexers when annotating records with indexing terms is commonplace. For many categories, however, the indexing terms nicely overlapped with the categories of the VODS. This gave us confidence in the construction. Since the expert validation yielded results comparable to previous exercises of this kind, we believe this procedure to be of sufficient quality to allow for an automated (re)classification experiment. On the other hand, one can also interpret the relatively low IIC scores as indicative of the inherent ambiguity at this level of granularity.

Future research and practical applications

The use of a minimum of textual data makes the approach presented in this study practical to generalize to other datasets, e.g. projects and project applications. Using additional bibliographic metadata would presumably increase the performance of the classification algorithms. Full-text documents would be an interesting path forward, yielding more textual data and a better sensitivity of TF-IDF transformations. In addition, it would be interesting to study ambiguities of the classification resulting from the predictions made by the algorithm and study those in detail.

With regard to the machine learning modules used, we acknowledge that more advanced and complex language processing techniques have a good track-record when it comes to automatically classifying text documents (e.g. BERT and related models). Dunham and colleagues (2020) have shown that SciBERT models outperform other NLP methods when applying them to classify publications in the field of Artificial Intelligence. For our purposes, however, we have opted to keep the set-up relatively straightforward. The main motivation behind this study was to investigate and compare the feasibility of using supervised machine learning algorithms for this particular, challenging fine-grained classification task. We leave comparisons of other methods and feature transformation procedures for future research.

Questions surrounding the properties of interdisciplinarity demand for a clear operationalization of disciplines, which is not straightforward. This is in itself also the main reason why many different classification schemes are used in different contexts, each pointing to insights about different aspects – organizational, cognitive, etc. – of a discipline (Guns et al., 2018). Textual approaches might lead to other insights regarding the cognitive structure of disciplines, but these same disciplines are in constant flux (Yan, Ding, Milojević, & Sugimoto, 2012). A fixed classification scheme will not meet future developments in science; "… human assigned subject categories are akin to using a rearview mirror to predict where a fast-moving car is heading" (Suominen & Toivanen, 2016, p. 2464). To this end, the team working on the VODS has provided a 'not elsewhere classified' category for all the subfields (Vancauwenbergh & Poelmans, 2019a). This particular category hasn't been studied in this article. Future deployments of the classification system in Flanders will allow researchers to identify themselves and/or their projects with this category and assign documents to it, and, following from this, we could study text residing in these categories to discover emerging research problems and topics.

Once researchers employed by the Flemish universities start to label their expertise, projects, and outputs using the VODS, supervised machine learning algorithms can be trained on a broader range of disciplinary categories, allowing for a broader evaluation of the method proposed in this paper. This approach will enable us in practice to assist with annotating unlabeled work, or it can serve to underpin an online recommendation system for researchers, embedded in current research information systems. The output of supervised text classifications can also be compared to other existing classification schemes. We can for example contrast the publication level classification with journal level classifications of the same publications in order to study the disciplinary or interdisciplinary diversity of journals.

Finally, we should highlight that measuring inter-indexer consistency is not straightforward. While there exists a long tradition of research which makes use of scoring systems like IIC or F1 scores to assess the reliability and functionality of classification systems, there have been attempts to include semantic relations between indexing terms or categories in order to develop more realistic measures of indexing accuracy. Medelyan and Witten (2006) propose to calculate the cosine similarity between word vectors of vocabularies or semantic definitions of categories. This is an interesting approach, but, to our knowledge, there are no systematic comparisons with other scoring systems available to date. It would be interesting to use such an approach when assessing classification systems in which a semantic definition of categories is available. The classification error could for example be weighted by the cosine distance between sentence embeddings of semantic definitions of the disciplinary categories. If a classification error is made whereby two distant categories are mistaken for each other, then the error is greater than when these categories are closer to each other in terms of cosine distance.

5.8 Conclusion

In this article we present a supervised machine learning approach to classify social science journal articles into multiple fine-grained disciplinary categories. Making use of Gradient Boosting with Classifier Chains we are capable of assigning one or more disciplinary categories to text documents (i.e., abstracts and titles). In order to do so, we have compiled a new dataset consisting of 113,909 records originating from three disciplinary databases in the social sciences (i.e., EconLit, ERIC, and Sociological Abstracts).

The novelty of this study lies in two aspects: (1) the construction of the labeled dataset, based on discipline-specific thesauri, and (2) the application of supervised machine learning algorithms to classify social science journal articles into one or more fine-grained disciplinary categories using text. We show in detail how we have collected the data and how we have validated the labeling based on the subject indexing terms from the thesauri. With regard to the machine learning methods, we compare different feature engineering techniques and two well-established classification algorithms. The Gradient Boosting classifier (LightGBM) in a Classifier Chaining framework is capable of predicting +/- 46% of the exact label combinations correctly, with a fraction of 0.3% of labels assigned incorrectly. The F1 score is 0.55.

In a previous study (Eykens et al., 2019) we assessed the performance of four different ML algorithms for the classification of sociology and anthropology journal articles extracted from Sociological Abstracts into fine-grained disciplinary categories (level 4 of the VODS). Making use of the same LightGBM module (Ke et al., 2017), we were able to correctly classify over 80% of the publications. In this previous study, we made use of simple feature engineering (i.e., lemmas and uni-grams) and we did not assess whether multi-label classification was possible. Aside from the work by Read (2010) to date, we are unaware of studies making use of similar methods to achieve fine-grained disciplinary classifications. To our knowledge, no work exists that studies the performance of supervised machine learning algorithms to classify social science documents on such a granular level.

Because we have significantly scaled up our dataset, this study adds more nuance to the previous experimental study (Eykens et al., 2019). We have added textual data from two additional disciplinary databases, namely ERIC and EconLit, and we have assessed more complex feature engineering techniques as well. Importantly, we assess whether multi-label classification is manageable. The results confirm the robustness of our previous work and expand it to additional data sources. We further demonstrate that to a certain extent the approach is indeed generalizable to a multi-label classification task. To achieve this, the quality of the data collection and data validation is crucial. Hence, we encourage others to develop a thought through data collection and validation procedure in order to make sure that the complete machine learning experiment is reproducible, from data collection and processing onwards.

To summarize, this study shows that supervised machine learning algorithms are capable of classifying social science journal articles into predefined, fine-grained categories based on the limited textual data of abstracts and titles only. However, for both human experts and machines such classification at the sub-disciplinary level proves very hard, to the extent that the question can be raised whether such an attempt makes sense. Given the need for fine-grained classification in view of assessments, evaluations and policy, we suggest that the informetric community further explores the possibilities for such fine-grained classification. For example, can the results obtained in this study be improved with different or more advanced NLP techniques, and by combining human expertise with advanced machine learning techniques? Like others (Boyack & Klavans, 2014; Suominen & Toivanen, 2016), we do not believe it to be fruitful to consider one or another classification system superior. We do instead insist that each approach has its merits, especially when contrasted to others. We hope that our work will spur others to conduct similar studies that explore the limits of the feasibility of classification through algorithms and human experts.

6 Assessing different document vectorization techniques for unsupervised clustering

Full reference: Eykens, J., Guns, R., & Engels, T. C. E. (2021*). Clustering social sciences and humanities publications: Can word and document embeddings improve cluster quality?* In W. Glänzel, S. Heeffer, P.-S. Chi, & R. Rousseau (Eds.), Proceedings of the 18th conference of the International Society for Scientometrics and Informetrics (pp. 369-374). Leuven, Belgium: International Society for Scientometrics and Informetrics.

Clustering scientific documents or publications based on relations between them has been at the core of scientometric research since the early years of the field (Small & Crane, 1979). When no or only very coarse document classification systems are in place, or when we want to develop a better understanding of the topics or specialties a document belongs to, clustering becomes an important go to method. It allows to organically construct groupings of similar documents based on specific commonalities between them from the bottom up (Waltman, Boyack, Colavazzi, & Van Eck, 2020). Depending on the granularity of the final clustering results, light can be shed on the knowledge base of disciplines, subdisciplines, or research specialties (Sjögårde & Ahlgren, 2018, 2019).

Throughout the years, many clustering techniques have been studied and the appropriateness of different bibliographic variables has been analysed. Co-citation relations between documents or journals and bibliographic coupling have traditionally been at the centre of the stage. In more recent years, topic modelling and other textual relations have been explored extensively (Boyack et al., 2011; S. Wang & Koopman, 2017). In addition, hybrid combinations of text and citation data have proven to be promising when it comes to identifying granular clusters of documents (Boyack & Klavans, 2014; Janssens et al., 2009).

In this study we investigate textual features and their usefulness for document clustering. We study the clustering outcomes for a dataset which contains publications from the social sciences (SS) and humanities (H) (SSH). For a considerable share of the documents in this dataset, no citation or reference data are available. Citation coverage is a well-known problem in the context of the SSH. We compare established document representation techniques such as TF-IDF and Latent Semantic Indexing with word and document embedding techniques (Word2Vec and Doc2Vec) in terms of the quality of the clustering outcomes. Quality is assessed by calculating silhouette scores, textual coherence of the clusters, and inspecting a cluster visualization. We conclude with a discussion of the results, the limitations, and pathways for future research.

6.1 Data

Data are collected from VABB-SHW, the Flemish bibliographic database for the SSH (Verleysen, Ghesquière, & Engels, 2014). We have selected all journal article publications for which abstracts and titles are available for the publication years 2011–2015. Titles and abstracts of English language publications have been merged together into one text field (TAs). TAs that were shorter than 65 words have been dropped; after this operation, 15,907 publications are left in the data set. SpaCy is used to tokenize the TAs. First a TA is split into individual tokens. These tokens are lower-cased and stemmed with the Porter stemming algorithm. Finally, stop words and punctuation are removed as well as tokens shorter than 3 characters.

6.2 Methods

k-Means clustering and clustering quality

In their seminal study 'Mapping the backbone of science', Boyack and colleagues (2005) make use of the k-Means algorithm to cluster documents based on their relative location on different maps. In more recent work, Wang and Koopman (2017) compare the performance of k-Means and the Louvain algorithm in the context of their Astro dataset. They apply these two clustering algorithms on semantic representations of the articles. For their application, both k-Means and the Louvain algorithm appeared to be competitive with other clustering solutions (Wang and Koopman, 2017, p. 1029). For this study, k-Means clustering is implemented in Python with the open source machine learning library scikit-learn (Pedregosa et al., 2011). The Mini Batch variant is used (Sculley, 2010). MiniBatch k-Means uses randomly sampled mini-batches or 'chunks' of the data to reduce the computation time. When a sample is drawn, the k-Means algorithm is run and the centroids are initiated and updated.

The quality of the clustering outcomes is measured by calculating average Silhouette scores and a metric for textual coherence. The Silhouette score is a measure of internal consistency or cohesion of the clusters compared to all other clusters (the separation) (Rousseeuw, 1987). The average Silhouette score thus gives an idea of the overall performance of the clustering solution. The textual coherence is a metric based on the Jensen-Shannon divergence (JSD). JSD computes the distance between two probability distributions; in this case the word vectors for the documents and the clusters they belong to. A higher score indicates more textually coherent clusters. This metric has been described extensively in Boyack et al. (2011).

For visual inspection, we map the clustering outcomes of k-Means on a t-SNE plot. t-SNE or t-distributed stochastic neighbour embedding is a highly effective technique well-suited for reducing high-dimensional data to two or three dimensions (van der Maaten & Hinton, 2008). Additionally, journal discipline classifications for the publications are mapped onto the t-SNE visualization.

Term frequency and inverse document frequency (TF and TF-IDF)

For each document, the number of times a term appears is counted (Term Frequency). Term frequency-inverse document frequency (TF-IDF) is a modification of this counting scheme and corrects for very common terms, or terms which are specific to a document. IDF is the inverse function of the number of documents in which that term occurs and is multiplied with the TF. This standard approach to text vectorization has proven itself over the years and has repeatedly been shown to be a worthy competitor for more advanced NLP (Natural Language Processing) techniques (Lelu & Cadot, 2019).

Latent Semantic Indexing (LSI)

The second approach tested is Latent Semantic Indexing (LSI), or Latent Semantic Analysis (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990). LSI makes use of Singular Value Decomposition to transform the document-term matrix into a reduced matrix with a fixed number of latent 'topics' or factors as features. LSI takes into account co-occurrences of words – the semantic meaning – to reduce a document to a predefined number of topics. The technique has been shown to be successful for large scale clustering applications with scientific publications (Boyack et al., 2011). We make use of the Gensim implementation of the LSI algorithm. We have evaluated the clustering outcomes for different numbers of latent topics ranging from 50 to 1000.

Word2Vec and Doc2Vec

Word2Vec is a technique to represent words with a unique list of numbers, the vectors (Mikolov, Chen, Corrado, & Dean, 2013; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). The context of the word is used to build a semantically meaningful vector representation of a word. It is a deep learning model that tries to predict the word to be embedded by training a neural network with the words in its direct context. Documents vary in length, however, and Word2Vec is a model that transforms individual words into vectors. Each document will be represented as a vector of n terms vectors. To obtain vectors of the same dimension for each document, we first sum all n terms word vectors per document and then calculate the arithmetic mean such that each document has a single x-dimensional vector of the same length. We use the Gensim implementation of the skip-gram model with negative sampling for our analysis.

Another extension of Word2Vec is Doc2Vec (Le & Mikolov, 2014). In Doc2Vec a document is represented by a vector of fixed size. Doc2Vec trains a model to construct a unique vector for every document. These unique document vectors are concatenated with the word vectors which are shared between documents and then averaged. Each document vector is thus a combination of a paragraph or document vector and Word2Vec vectors. Doc2Vec has been implemented in the study of scientific publications and has, for example, been used in conducting detailed similarity analysis among paragraphs (Thijs, 2020).

6.3 Results

The optimization of the k value for k-Means according to the average silhouette scores led to very different results in terms of the granularity of the clustering solution. TF-IDF and LSI produced solutions of 73 and 42 clusters respectively. LSI produces clusters of higher quality in terms of silhouette scores. In contrast, the textual coherence is higher for TF-IDF. This might not be a surprise, as smaller clusters tend to be more textually coherent.

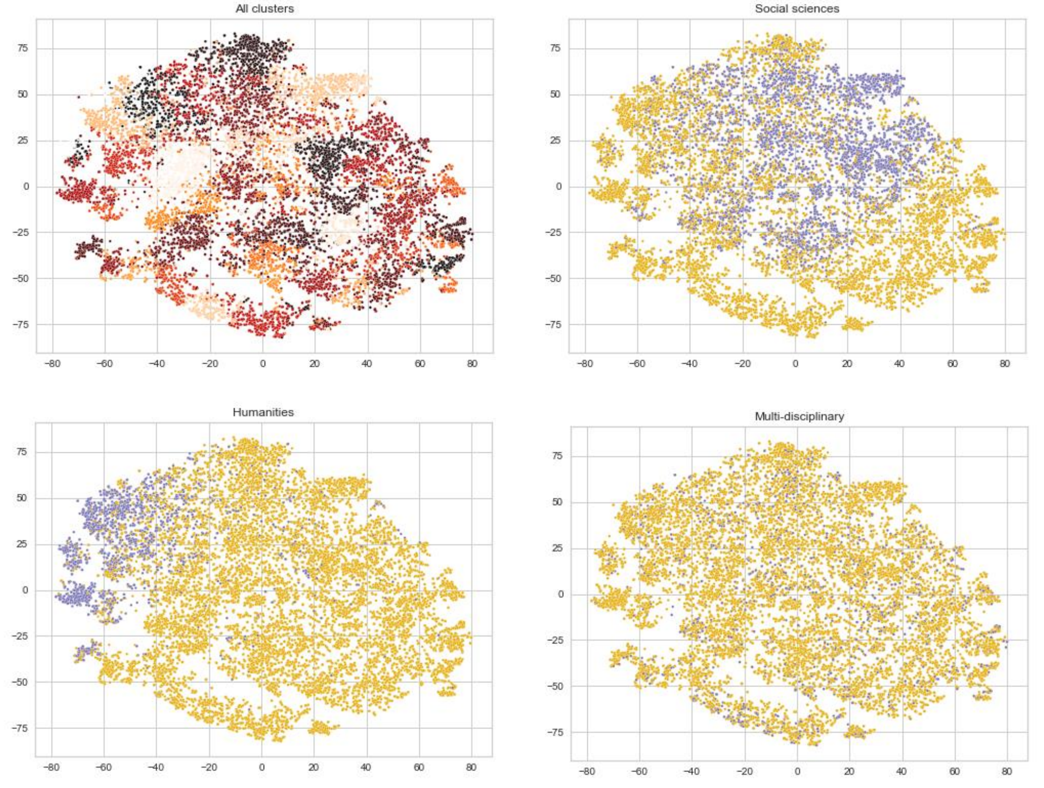
	Number of clusters	**Feature vector size**	**Average Silhouette score**	**Textual coherence (above random)**
TF-IDF	73	9,849	0.01362	0.0422
LSI	42	50	0.0724	0.0389
Word2Vec average	40	25	0.0908	0.0434
Doc2Vec	22	15	0.0520	0.0324

Table 7. Clustering quality for the different feature processing techniques.

The Word2Vec average strategy turned out to be the best performing feature vectorization method. With this processing technique, the k-Means algorithm divided the set of documents into 40 clusters. Whereas the average silhouette score is only slightly better than that for LSI, the average textual coherence is higher than was the case for other methods. This is an interesting finding, as the textual coherence is biased towards smaller clusters. As already noted by (Boyack et al., 2011), here too the textual coherence of the clusters decreases as the size increases. It would thus be no surprise if we were to find a higher coherence for the TF-

IDF solution. The latter yielded 73 and generally smaller clusters. The computational cost, however, is quite large. The number of features for TF-IDF is 394 times larger than the number of features needed for Word2Vec.

Figure 9. Visualization of t-SNE embedding. (a) upper left: K-means overlay, (b) upper right: publications in SS journals highlighted, (c) publications in H journals highlighted, (d) publications in multi-disciplinary journals highlighted.



In figure 9 we display t-SNE visualizations of the Word2Vec average document embeddings. The scatterplot in the upper left corner overlays the t-SNE embedding with shades indicating the clustering generated by k-Means. It becomes clear that, although most groups are more or less distinguishable in both cases, there exists considerable overlap between the clusters generated by k-Means. The remaining three plots overlay the t-SNE embedding with journal classifications of the publications. The journals have been classified according to the OECD fields of science (for details see Guns et al.

(2018). In the upper-right corner, publications in SS journals are highlighted. We can see that the upper-right section of the scatter plot is densely populated with publications in SS journals, gradually fading to the upper left.

On the scatterplot in the lower-left corner publications in humanities journals are highlighted. Here we see that they are positioned opposite to the publications in SS journals highlighted in the upper-right scatterplot. The upper left region of the scatterplot highlighting the humanities seems to be most densely populated, fading to the right hand side, the corner in which SS publications are. This fading over is an interesting pattern, indicating that there might be important cognitive cross-over areas between both the SS and H. As we can see in the upper left corner, one of the clusters which is found by k-Means (in white) nicely overlaps this fading area. The fourth and last scatter plot highlights publications in multi-disciplinary journals. These are journals which have been assigned to multiple OECD FOS discipline codes. As one would expect, these publications are dispersed over the scatterplot, popping up in multiple clusters of documents.

6.4 Limitations

The clustering algorithm used in this study demands for a pre-specified number of clusters. Additionally, only one similarity measure has been tested (Euclidean distance). It would be interesting to compare the k-Means clustering algorithm as well as the similarity measure to other current approaches used in a bibliometric context. Boyack et al. (2011), for example, make use of the DBSCAN algorithm and cosine similarity. Other studies make use of the Leiden algorithm or the Louvain algorithm for community detection (Wang & Koopman, 2017). Similarity approaches like the cosine similarity and BM25 similarity have been shown to perform well in a text clustering context (Waltman et al., 2020).

6.5 Conclusion and future research

In this paper we have shown that state-of-the-art vectorization techniques (Word2Vec and Doc2Vec) work well in the context of clustering SSH publications based on textual information. Although the outcomes of the different vectorization techniques in terms of silhouette scores and textual coherence were nearly identical, Word2Vec average turned out to be the best strategy to identify 40 well-divided and coherent clusters. The contextual sensitivity of Word2Vec might be a possible explanation.

A next step consists of studying different clustering algorithms and similarity calculations. Additionally, studying these clusters of documents in detail will yield interesting insights into specialized communities. The clusters can be thought of as the knowledge base of research specialties or disciplines. In another phase we will add an additional, more granular layer to the clustering presented above. A detailed content analysis and second round of clustering for the different document groups will yield insight into what these clusters actually represent. Are they representations of one or many subject specialties?

We hypothesize that while some document clusters might be representative of local or regionally oriented specialties, others will be part of the knowledge base of more global communities. For the latter, it will be difficult to reach any conclusions without taking the broader context into account. For the former, however, by studying their bibliographic, cognitive and social characteristics we will be able to broaden our understanding of research specialties in the SSH as well as bibliographic units.

Part 3 – The interplay between and isolation of disciplines in the social sciences and humanities

7 Crossing disciplinary boundaries? Cognitive and disciplinary mobility in the social sciences and humanities

Full reference: Eykens, J., Guns, R., & Engels, T. C. E. (under review). Crossing disciplinary boundaries? Cognitive and disciplinary mobility in the social sciences and humanities.

In debates about science, disciplines are often regarded as silos or insulated structures imposing boundaries on knowledge exchange and innovation (Jacobs, 2013). These assumed boundaries however are contested; they are constantly being (re)negotiated and subject to change (Massey, 1999). It has for example been proven hard to achieve consensus on a possible definition of the notion of 'discipline' (Sugimoto & Weingart, 2015) or indicate what the core set of journals for a discipline are (Sīle et al., 2021). One of the factors contributing to this dynamism is differentiation of the scientific system into sub-disciplinary specialties further enhancing potential interdisciplinary interactions across disciplines (Stichweh, 1992, 2003). A recent resurgence in the interest for interdisciplinary has sparked further research into the *disciplinary silos* hypothesis, albeit in the form of a nuanced discussion on different aspects of knowledge transfer (where and when knowledge transfer takes place between different disciplines or fields, for example) (Larivière & Gingras, 2014).

The largest share of the (mainly scientometric) literature on knowledge flow or diffusion across fields of science is based on citation network analysis (Liu, Hu, & Li, 2018; Liu, Shi, & Li, 2017; Lockett & McWilliams, 2005; Rinia, van Leeuwen, Bruins, van Vuren, & van Raan, 2002; Yan, 2014; Yan & Ding, 2012). Exporting or importing knowledge to and from disciplines, however, can effectively take place when an author carries her specific experience or knowledge from one field to another (Chubin, 1976; Hargens, 1986). A separate strand of scholarship therefore investigates knowledge diffusion by analyzing field mobility of authors (Chakraborty, Tammana, Ganguly, & Mukherjee, 2015; Ferreira & Costas, 2021; Hargens, 1986; Le Pair, 1980; Urata, 1990; van den Besselaar, 2019b; van Houten, van Vuren, Le Pair, & Dijkhuis, 1983). How do scientists move from one discipline or topic to another throughout their careers? And what are the potential effects?

Each discipline may play a unique role in such a knowledge diffusion network. Medicine, physics and chemistry act as 'sources' or knowledge exporters, while more applied fields, such as chemical engineering, material science, or neuroscience behave like 'sinks' or knowledge importers (De Domenico, Omodei, & Arenas, 2016). For disciplines from the social sciences and humanities research (Le Pair, 1980) shows that classical languages and literature, geography, and history serve as 'donors'. Medicine (as opposed to what (De Domenico et al., 2016) found), agricultural sciences, theology, non-western languages, archeology, arts, philosophy and social and cultural sciences mainly function as 'receptors'. The same study (Le Pair, 1980) further finds that scholars from the professorial ranks are less mobile than their junior counterparts. Researchers are 'never too old to migrate' (p. 260) but they are most mobile in the early part of their careers.

The relevance of changes in an authors' research direction is not limited to the potential of observing exchange of information between fields. It has for example been shown that those who switch between disciplines or sub-disciplines increase their scientific impact. Scholars from physics and astronomy who change their orientation and travel further away from their root topics receive more citations on average than their counterparts who do not move far (Yu, Szymanski, & Jia, 2021).

For fields like the social sciences and humanities field mobility can be used as a complement to citation analysis (Urata, 1990). Urata studied field mobility patterns for social sciences and humanities scholars from Japan and compared the mobility network to a citation network. The findings indicated that the two approaches yield a similar structural picture. Methodologically speaking, using field mobility by authors might thus offer the potential to gain fresh insights into field interdependence and interdisciplinarity in the SSH, be it by following a somewhat different logic than, but related to citation relations (e.g., knowledge diffusion through publications as apposed to diffusion through citations, see Liu & Rousseau, 2010).

To summarize, discipline switching or mobility by individual researchers might teach us something different about the overall mutual dependence between disciplines. Do authors mainly move between disciplines which are cognitively similar? Do they often switch between disciplines? On the level of research specialties, it has for example been argued that different types of cognitive 'migrants' exist (Chubin, 1976, p. 468). Those who do not travel, or work in one or two closely related specialties throughout their career, researchers who frequently travel between specialties, but do not necessarily travel far, and those who travel frequently and far. Researchers from the latter category share the characteristic of 'intellectual breadth'. Similar to what (Chubin, 1976) further argues for specialties, we might find such patterns on the level of disciplines as well, e.g. groups of researchers who visit many disciplines throughout their career and travel far. As such, the findings might add nuance to the common idea of a stable disciplinary system as well as the permeability of disciplinary boundaries.

The literature on disciplinary mobility, however, remains scattered. Dating back to the 1980s we were able to identify 26 quantitative studies which were more or less related to the topic. While yielding additional potential for scientometric studies of the social sciences and humanities, even fewer bibliometric evidence is available for these fields. We aim to fill this gap by studying disciplinary and cognitive mobility in the SSH based on comprehensive bibliographic data for Flemish researchers from the social sciences and humanities.

Aims and outline of the study

Urban (1982) defined cognitive migration as a 'change of cognitive and socio-cognitive situations of production, with or without geographical social mobility' (p. 412). Cognitive situations are written texts and discourse within the social system of interaction and communication. The journals (or other media) in which a researcher publishes her results can be understood as such a socio-cognitive situation of production. In the first part of this study, we draw an overall picture of disciplinary switching patterns for SSH researchers. We do this based on two network representations. A first network is derived from yearly changes of the main (the mode) disciplinary classification of authors' publications. We discuss some basic characteristics of the networks, like the centrality of disciplines, the relative closeness of disciplines to one another in terms of switching flow (shortest paths) and if clusters are present. For example: are the social sciences and humanities two separate subgroups in the networks?

We also elaborate on characteristics of the individual disciplines and their position relative to each other. Which disciplines can be regarded as knowledge exporters or absorbers (ratio of in- and outflow of researchers)? And what is the migration breadth, i.e., the 'degree' in network terminology (number of other disciplines to which a discipline is connected)? The second network represents the cognitive similarity between disciplines. We put these networks next to each other to investigate structural (dis)similarities and answer the question whether authors mainly switch between disciplines which are cognitively close to each other.

In the second part of the results section, we study the distance traveled by scholars in terms of a continuous measure, i.e. the cognitive similarity (or cosine distance) of the contents of their work (the methods used are discussed in more detail below). Besides a general question about career stage or 'academic age' and cognitive mobility, we would like to understand to what extent disciplinary mobility as exemplified by publishing in different discipline categories is related to the actual cognitive distance traveled by an author. Do authors move far when they cross disciplinary boundaries or are they working on the intersections between different disciplines (i.e., interdisciplinary specialties)? Metaphorically speaking: do authors work near the border, or do they travel abroad?

To summarize the main research questions discussed in this study:

(i) *Where does knowledge diffusion occur for the SSH through authors switching between disciplines? What are the roles played by different disciplines?*
(ii) *Do authors mainly switch between fields that are cognitively similar?*
(iii) *Are junior researchers more mobile on average than their senior counterparts? Are researchers more cognitively mobile in their early careers?*
(iv) *Do authors travel far when they cross many disciplinary borders, or do they conduct boundary work?*

In the next section we will first present the dataset which is used for our analysis. Subsequently we discuss the network analytical approaches used and the operationalization of cognitive similarity and mobility. The results section follows the same ordering as the outline of the research aims presented above. We first look at the structural characteristics of the discipline switching network. Second, we present our analyses of the cognitive similarity between disciplines and the distance traveled by the authors. This is followed by a discussion of the implications of our findings, and the main limitations of our analyses. We conclude by presenting potential avenues for future research.

7.1 Data

We study author mobility based on data from a local and comprehensive bibliographic database specifically designed for better coverage of research output produced by researchers from the social sciences and humanities in Flanders, Belgium (the Dutch speaking part of the country). The dataset covers all publications (168,641) authored by scholars who are (or were) affiliated to a SSH research (18,422) unit at one of the five Flemish universities, and that appeared between 2000 and 2018. The database is multilingual and covers multiple publications types, including journal articles (105,681 or 63 %), books (5,588 or 3.3 %), book chapters (43,221 or 25.6%), edited volumes (6,740), and conference proceedings papers (7,411) (Engels & Guns, 2018; Verleysen et al., 2014). 116,854 (±70%) records are peer reviewed and 47,611 are not.

Two disciplinary classification systems are provided in VABB-SHW. One is based on the affiliation address of an author, and the other on the cognitive scope of a publication or the journal/source in which a publication appeared. The latter will be used for our analyses, as it is available for a large share of authors. For the authors, we face the problem that affiliation data are only available for those authors in a byline who have been active at a Flemish SSH department or research group. The cognitive classification is a slightly adjusted version of the OECD Fields of Science classification system (described in detail in: Guns et al., 2018). OECD FOS contains 45 discipline categories (OECD, 2007), ranging from natural sciences to social sciences and the humanities.

The following SSH discipline categories are present in the OECD Fields of Science classification system. In VABB-SHW categories 6.1, 6.2, and 6.3 are further subdivided into their constituent fields:

> *5. Social sciences:* 5.1 Psychology - 5.2 Economics and business - 5.3 Educational sciences - 5.4 Sociology - 5.5 Law - 5.6 Political science - 5.7 Social and economic geography - 5.8 Media and communications - 5.9 Other social sciences
>
> *6. Humanities:* 6.1.1 History - 6.1.2 Archaeology - 6.2.1 Languages - 6.2.2 Literature - 6.3.1 Philosophy and ethics - 6.3.2 Religion - 6.4 Arts (arts, history of arts, performing arts, music) - 6.5 Other humanities

Here we should note that, as a result of co-authorship, collaboration and multiple affiliations across disciplines and specialties, publications in other fields than the SSH are included in the dataset as well. In the analyses, because of their partial coverage, we only pay scant attention to fields other than the SSH, i.e., if they are in some way found to be related to the latter.

For the creation of the dataset, we select those records for which OECD FOS codes are available. We further reduce this set to publications of which the language is covered by the embedding model used (discussed in the next paragraphs). These languages are: English (64%), Dutch (27 %), French (5 %), German (2%), Spanish, Italian, Russian, Chinese, Portuguese, Polish, Japanese, Turkish and Arabic (all < 1%).

7.2 Methods

7.2.1 Discipline switching network

For the construction of the author mobility network, we want to take into account all unique mobility events for each author who has at least 4 years of activity registered in VABB-SHW. 4 years is the typical duration of a PhD trajectory and might thus be an adequate threshold for retaining those researchers with a career in academia. For an author who has published in 4 or more publication years, we select the modes of the discipline codes for every publication year. To establish edges in the network, we pair all discipline codes at time *t* with those from *t-1* and fractionalize the counts when multiple modes are present (by dividing 1 by the total number of modes at time t + at t-1). Non-moves are not counted.

Year	Modes
2000	6.2.2; 6.1.1
2001	5.3
2002	6.3.1; 5.3 ; 5.2
2003	6.2.2

period	source	target	weight 1/(D*t* + D*t-1*)
2000-2001	6.2.2	5.3	0,33
2000-2001	6.1.1	5.3	0,33
2001-2002	5.3	6.3.1	0,25
~~2001-2002~~	~~5.3~~	~~5.3~~	~~0,25~~
2001-2002	5.3	5.2	0,25
2002-2003	5.3	6.2.2	0,25
2002-2003	6.3.1	6.2.2	0,25
2002-2003	5.2	3.2.2	0,25

source	target	weight
6.2.2	5.3	0,33
6.1.1	5.3	0,33
5.3	6.3.1	0,25
5.3	5.2	0,25
5.3	6.2.2	0,25
6.3.1	6.2.2	0,25
5.2	6.2.2	0,25

Table 8. Creating the disciplinary mobility edgelist.

We construct a directed network, meaning that switches from *discipline A -> discipline B* and *from discipline B -> discipline A* are considered different directions. In the case of duplicate switches on the level of an author, we only keep the move with the highest weight. After collecting all switching pairs for each author, we deduplicate them based on author identifiers and their direction. All weights obtained for the individual authors are summed. This aggregate information results in an edge list of source and target disciplines which is turned into a weighted and directed network. An illustration of the creation of an edge list for a hypothetical author is presented in table 8.

7.2.2 Cognitive similarity network and mobility

Cognitive similarity and mobility are studied by analyzing changes in the scope or cognitive content of publications, on the level of individual authors and on the level of disciplines. To construct a discipline similarity matrix, for each category we collect the titles of the publications with only one discipline category assigned (28,256 documents have multiple disciplines assigned or about 17%). Document titles are vectorized with a pre-trained Universal Sentence Encoders model (Universal-sentence-encoder-multilingual, obtained from https://tfhub.dev/google/universal-sentence-encoder-multilingual/3). This model is able to handle 16 languages and has been optimized for short text vectorization (sentences, paragraphs, or short paragraphs).

The result is a 512-dimensional vector for each title record. We compute the arithmetic mean of the document embeddings for each discipline category to obtain a discipline vector taken as representative for the overall cognitive contents of that discipline. This thus results in a matrix representing a single 512-dimensional embedding per discipline category. Next, we calculate the cosine similarity between all pairs of discipline vectors to obtain a similarity matrix of disciplines. We also compute the averages of the vectors of titles obtained for each author and every publication year in which she has been active. To study the cognitive distance traveled from year to year, we calculate the cosine distance between these vectors for the different publication years. We have retained records with very short titles (< 50 characters) from the analyses, as they introduced quite some noise in the distance calculations. We have also checked the robustness of the results with a subsample of (6,466 or 83% of the full sample) authors with records for which full abstracts were available, but they did not seem to differ in any noticeable way from the result obtained with titles only.

For the nearest neighbors analysis, the weighted cognitive similarity network without weak edges is used, and for each discipline the three nodes to which it is most strongly connected are retrieved, both from the cognitive similarity network and the disciplinary mobility network. The disciplinary mobility network is directed, so we differentiate between predecessors and successors, i.e., disciplines from which researchers migrate, and disciplines to which researchers emigrate respectively.

For an overall structural comparison of the networks, we first calculate the cosine similarity between the two networks (Yan & Ding, 2012). The matrices of the two networks are transformed into two 1-dimensional vectors or ‘flattened’ and then the cosine similarity between these vectors is calculated. For centrality measurement, we calculate PageRank centrality in NetworkX (Hagbert, Schult, & Swart, 2008). PageRank takes into account the weights of the incoming edges of a node. Clustering is carried out in VOSviewer (van Eck & Waltman, 2010). The software makes use of a modularity-based clustering algorithm as proposed by (Clauset, Newman, & Moore, 2004), capable of handling weighted networks. Directed networks are treated as undirected ones, meaning that an edge from *A* to *B* is treated the same as one in the opposite direction.

7.2.3 Mobility and broadening occurrences/shifts

Disciplinary mobility and broadening occurrences are operationalized by studying the main discipline categories for an author in each publication year and changes therein. We distinguish between two types of mobility events: a general disciplinary mobility event and a more specific disciplinary broadening event. As described above, for each author we have the mode disciplinary category for each year in which she has published available. In some cases, multiple modes are retrieved. If so, we check for intersections of these sets. Next, to indicate a mobility event, it is checked whether the mode (or set of modes) at time *t* differs from the mode of *t-1*. If there is overlap, we indicate the mobility event column with 'false', and 'true' in the case of no overlap. *t-1* can be considered a shift back in a standardized timeframe, meaning that two publication years do not need to be adjacent in time. If an author is inactive for one or two years, these are dropped from the analyses and the shift from, say, 2000 to 2003 is taken into consideration.

For the disciplinary broadening events, it is checked whether a new discipline is visited at time *t*. This is done by checking whether a mode at time *t* is a new discipline for an author. Formally we check if a mode at time *t* is already present in one of the previous publication years. Naturally the first year of each author is indicated with 'false' for both variables. Table 9 contains an illustration of the concepts of disciplinary mobility and broadening. If no overlap exists, as is the case for 2000-2001, a mobility event has occurred and is indicated as such for year 2001 *t*. If overlap exists, as is the case for 2002-2003, no mobility event is counted for year *t*. The broadening events indicate if for a specific publication year, a new discipline occurs and if an author 'broadens' her disciplinary horizon.

Year	Modes	Mobility event	Broadening event
2000	6.2.2; 6.1.1	N/A	N/A
2001	5.3	1	1
2002	6.3.1; 6.2.2	1	1
2003	6.2.2	0	0
2004	6.3.1	1	0
2005	5.5; 6.3.1; 6.1.1; 6.2.1	0	1
2006	6.2.2	1	0
2007	6.2.2	0	0
2008	6.3.1	1	0

Table 9. Counting mobility and broadening occurrences.

7.3 Results

7.3.1 Comparing the discipline switching and the cognitive similarity networks

In table 10 we list the descriptive statistics for both networks, before and after removing weak edges. The initial discipline switching network consists of 45 nodes with 1,426 weighted and directed edges. The median and average edge weights are 1 and 9.32 respectively, with a standard deviation of 26.79. To remove noise (in terms of what can be considered insignificant discipline switches) and reduce the initial density of the network (0.72) all edges with a weight below 2 are removed. These are edges representing the switches of 2 or less researchers. The resulting network consists of 819 edges (with a density of 0.41).

	Type	**Number of connected nodes**	**Number of edges before → after removing weak edges**	**Density before → after removing weak edges**	**Number of clusters**	**Size of largest cluster**
Discipline switching network	Weighted and directed network	42	1,426 → 819	0.72 -> 0.41	4	19
Discipline similarity network	Weighted an undirected network	43	946 → 467	1 -> 0.5	3	17

Table 10. Descriptive statistics discipline switching and discipline similarity network.

The initial weighted and undirected discipline similarity network is fully connected (density = 1), with 946 edges and a median and average edge weight of 0.65 and 0.63 (standard deviation = 0.16). We remove insignificant similarities (below 0.65) to allow for more appropriate clustering outcomes. Edges with a weight below 0.65 are set to 0, and this results in a remaining 467 edges (density of 0.494). For this network, Industrial biotechnology and Nano-technology are not included. This is because for Nano-technology no publication has this discipline assigned to it as its first and only category. Industrial biotechnology was only weakly connected. Hence, only 43 nodes are present within this network.

Global and local structure

How are the disciplines related to each other - in and across the different networks? We first compare the global similarity and clustering results for the two networks to study if their overall structure is similar. Are cognitively similar clusters of disciplines also clustered together in the disciplinary mobility network? Next, we study the local structure (on the level of individual disciplines) by investigating overlap between lists of nearest neighbors. The question which we would like to answer here: do researchers indeed mainly move between disciplines which are cognitively similar?

Network similarity and node centrality

The cosine similarity (0.42) between the two networks indicates that they are not entirely dissimilar. (Yan & Ding, 2012) for example, find very high cosine distance scores (0.9+) when they compare citation and co-word or topic networks, but this does not seem to be the case here. The most central nodes in the discipline switching network according to the weighted PageRank centrality: 5.4 Sociology (0.076), and 5.2 Economics and business (0.072), and Clinical medicine (0.064). Ranked by degree centrality, Economics and business lists first, followed by Sociology, History, and Art. For the cognitive similarity network, the most central disciplines by degree are: Sociology (34 edges), Social and economic geography (34 edges), and Other social sciences, (34 edges).

Clustering outcomes

For the clustering of the cognitive similarity network, we have first dichotomized the edge weights. Edges with a weight below 0.80

(average + 1 standard deviation) are weighted with 1, edges with a weight between 0.80 and below 0.95 are weighted with 2, and edges with a weight above 0.95 are weighted with a value of 3. Industrial biotechnology and Nano-technology are isolates, so there are 43 nodes in the network.

Cluster 1 in the network (right hand side figure 10) consists of all social sciences and humanities disciplines and Other medical sciences, except for Psychology and Economics and business. Also, the category Other natural sciences is present within this cluster. In the visualization we see that Social and economic geography and Sociology are pulled closer to the side of natural and engineering sciences. Political science, Art, and Languages seem to bridge the social sciences and humanities.

Cluster 2 consists of almost all natural and medical science disciplines, including Veterinary science, Psychology, and excluding Earth and related environmental sciences. Cluster 3 lastly, contains the latter together with engineering sciences and technologies, and agriculture, forestry and fisheries disciplines.

The discipline switching network consists of 4 clusters (figure 11). Nano-technology, Health biotechnology and Agricultural biotechnology are isolates. The first cluster contains all disciplines from the natural sciences but Other natural sciences, together with the engineering, agriculture, forestry and fisheries disciplines, and Economics and business and Social and economic geography from the social sciences group. Cluster 2 consists of all humanities disciplines together with Other natural sciences. Cluster 3 contains all medical sciences together with Psychology, Medical engineering, Veterinary science, and Industrial biotechnology. Cluster 4 consists of Educational sciences, Sociology, Law, Political science, Media and communications, and Other social sciences.

The structure of the two networks looks similar in the sense that social sciences and humanities disciplines are in general clustered together. Psychology and Economics are exceptions in this regard. Psychology is mostly clustered together with medical fields. Economics and business is grouped together with natural science disciplines in the cognitive similarity network. In the discipline switching network Economics and

Figure 10. Cognitive similarity network (3 clusters). An interactive version of this graph can be consulted online:

https://tinyurl.com/y8w2a59k

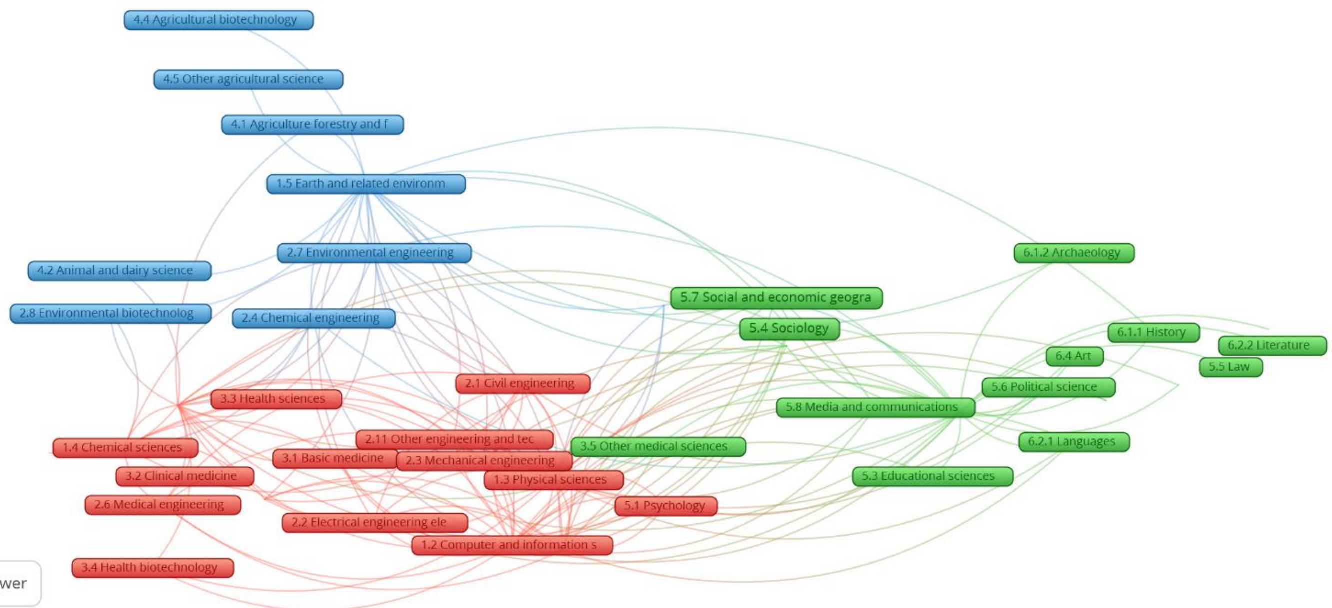
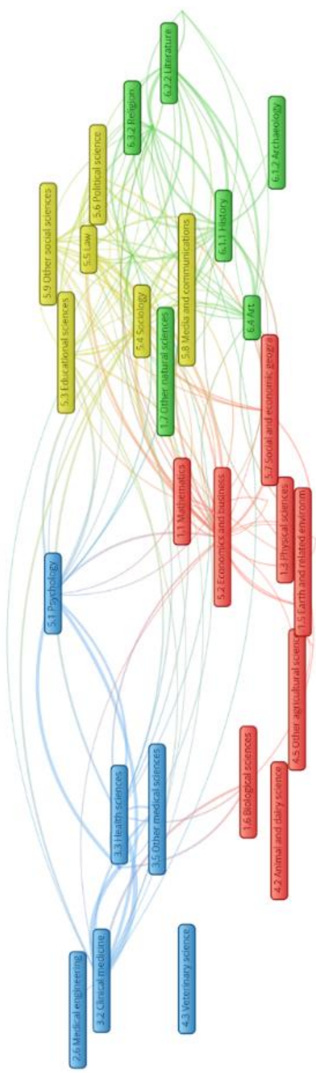

4.4 Agricultural biotechnology
4.5 Other agricultural science
4.1 Agriculture forestry and f
1.5 Earth and related environm
2.7 Environmental engineering
4.2 Animal and dairy science
2.8 Environmental biotechnolog
2.4 Chemical engineering
6.1.2 Archaeology
5.7 Social and economic geogra
5.4 Sociology
6.1.1 History
6.2.2 Literature
6.4 Art
5.5 Law
5.6 Political science
2.1 Civil engineering
3.3 Health sciences
5.8 Media and communications
2.11 Other engineering and tec
3.5 Other medical sciences
6.2.1 Languages
1.4 Chemical sciences
3.1 Basic medicine
2.3 Mechanical engineering
3.2 Clinical medicine
1.3 Physical sciences
5.3 Educational sciences
2.6 Medical engineering
5.1 Psychology
2.2 Electrical engineering ele
1.2 Computer and information s
3.4 Health biotechnology
wer

Figure 11. Disciplinary switching network (4 clusters). An interactive version of this graph can be consulted online: https://tinyurl.com/yce84rot


6.2.2 Literature
6.3.2 Religion
6.1.2 Archaeology
5.6 Political science
5.9 Other social sciences
5.5 Law
6.1.1 History
5.8 Media and communications
5.4 Sociology
1.7 Other natural sciences
6.4 Art
5.3 Educational sciences
5.7 Social and economic geogra
1.1 Mathematics
5.2 Economics and business
1.3 Physical sciences
1.5 Earth and related environm
5.1 Psychology
4.5 Other agricultural scienc
1.6 Biological sciences
3.3 Health sciences
3.5 Other medical sciences
4.2 Animal and dairy science
3.2 Clinical medicine
2.6 Medical engineering
4.3 Veterinary science

Business and Social and economic geography are clustered with natural sciences and engineering disciplines. We can already conclude that a comparison of the clusters in these networks partially contradicts the hypothesis that researchers mainly move between disciplines which are cognitively similar. Let us now take a look at the local mobility and similarity environments of each of the disciplines.

Nearest neighbors analysis

Mathematics has a strong mobility connection to Business and economics and Psychology, which could be interpreted as an indication of the ongoing mathematization of these disciplines (Katzner, 2003). Computer and information sciences has a strong mobility connection to SSH disciplines like Media and communications, Business and economics, and Languages, which might be due to the turn toward computational models (digital humanities and computational linguistics) and big data analyses in SSH scholarship (Conte et al., 2012; Wyatt, 2016). An established tradition in scientometrics is another possible explanation for these mobility connections. Medical sciences disciplines show an overlap of two disciplines on average. For Basic medicine, Clinical medicine, and Health sciences, Psychology is a successor and predecessor in the mobility network, but is not listed in the top 3 most similar disciplines. A perfect overlap is found for Other medical sciences.

5 out of 9 social sciences disciplines do not have a strong mobility connection to the cognitively most similar disciplines. Other social sciences, Media and communications, Political science, Sociology, and Business and economics have only one overlapping discipline. For Psychology, Educational sciences, Law and Social and Economic geography, we find two overlapping disciplines. No discipline has a perfect overlap of top 3. Psychology is in many regards strongly connected to medical fields. Sociology is listed in its similarity environment, and Educational sciences is a predecessor.

There is also a lot of mobility between Sociology and Business and economics, but Economics and Business is not listed in Sociology’s top 3 cognitively most similar disciplines. 7 out of 9 disciplines have Economics and business and Sociology in their immediate mobility environments. These are remarkable findings. On the one hand, previous evidence for the isolationist nature of Economics is contradicted by these results (Truc, Santerre, Gingras, & Claveau, 2020). Sociology on the other is again found to be interstitial (Abbott, 2001; Small & Crane, 1979). Sociology and Educational sciences function as mobility bridges between SS and the H.

Disciplines in the humanities have a fairly strong mobility connection to disciplines which are cognitively most similar (2 out of 3 disciplines on average), with History, Languages, and Philosophy and ethics being the exceptions. For the first two only one discipline overlaps. For Philosophy and ethics, no overlapping disciplines are found. This discipline is most similar to Other natural sciences, Other humanities, and Other social sciences, but mobility patterns are mainly established between Religion, History, and Sociology. History and Literature are present in the mobility environment of all Humanities disciplines.

While we can conclude that the discipline switching network is different from the discipline similarity network on the local level, and we also find that in both representations, the social sciences and humanities spheres are fairly disconnected from each other.

Importers vs. exporters

Comparing the in- and out-degree values for each discipline category allows us to study if disciplines can be considered knowledge importers or exporters. When contrasting the two values, we obtain a ratio of in- vs. outflow of researchers. Interestingly, for every discipline this ratio is close to 1 (no difference between the number of outgoing and incoming researchers). Exceptions of disciplines with a slightly higher number of outgoing researchers (ratio < 1) are Archaeology, Religion, Literature, Other humanities, Economics and Business, Law, and Other social sciences. Disciplines which could be considered importers (ratio > 1) Educational sciences, Political science, Social and economic geography, and Media and communications.

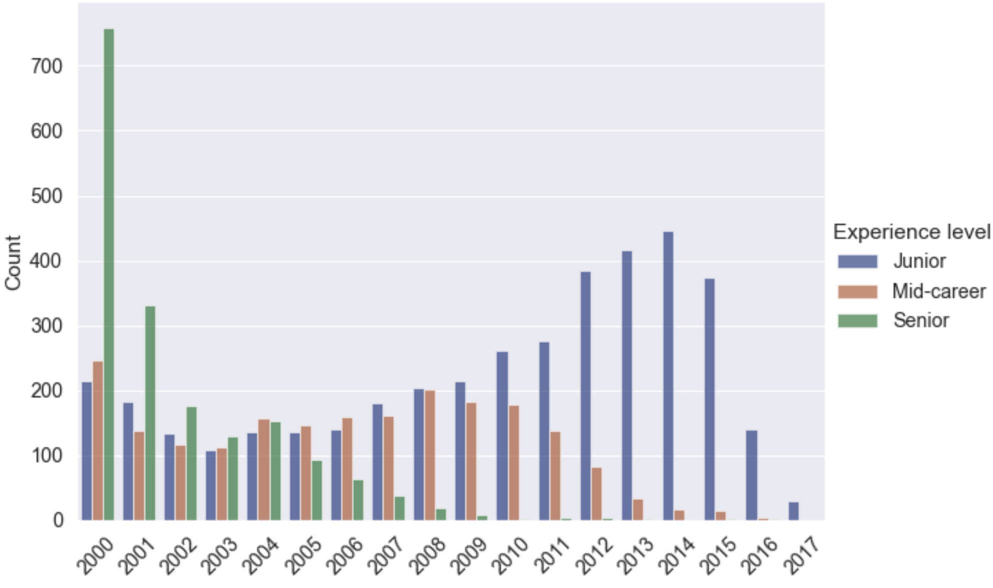
7.3.2 Disciplinary boundary crossing and cognitive distance traveled

Individual researchers are the main actors within the knowledge diffusion network. The act and intensity of diffusing knowledge from one domain to another can be approximated by studying the number of times authors cross disciplinary boundaries (or in this study, switches between the main disciplinary categories in which she publishes). The following analyses are conducted on a subset of researchers with four or more years of activity registered (7,829). Records with titles of less than 50 characters have been dropped, leaving 44,773 switching events in the sample. It should not come as too big of a surprise that the number of years of activity is positively correlated with the number of disciplinary switching or broadening events. Because of this, we divide the set of authors into three categories:

- Junior researchers (4 – 7 years of activity registered): 3,969
- Mid-career (8-12 years of activity registered): 2,081
- Senior (13+ years of activity registered): 1,779

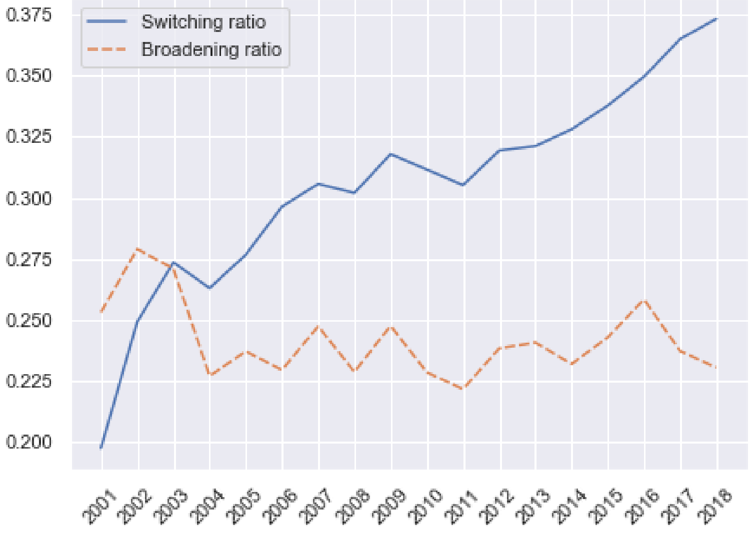
Figure 12 shows the distribution of the number of active authors per first year of activity divided over the career stage groups. Up until 2003, we see that the largest share of active authors is senior. As a consequence of our operationalization, this group is naturally more spread out over the first publication years. Additionally, in Flanders, the overall number of junior and post-doc researchers has increased more than the number of senior staff over the past 15 years. This overall growth has also flattened out between 2010-2012 and again between 2015-2017 (https://www.vlaamsindicatorenboek.be/3.3.1/evolutie-van-het-aantal-onderzoekers). We can also see that some senior and mid-career researchers are present in later years. This has to do with the cleaning of short titles. In these cases, publications of later years are possibly considered as their first year of activity.

Figure 12. Number of authors grouped by their first year of activity and career stage. Junior researchers (4 – 7 years of activity), mid-career (8-12 years of activity), and senior (13+ years of activity).



The vast majority of researchers in the sample are disciplinary mobile. Only 22% of authors has not broadened their disciplinary horizon and only one out of three (32.7%) did not switch between disciplines. The number of discipline categories in which the authors are able to publish remains the same over the period analyzed. Nonetheless, as explained in the introduction, we expect that authors on average switch more between different disciplines categories over time. Due to an increased specialization, it is expected that sub-disciplinary and specialized research groups are further established at the intersections of different disciplines, resulting in an overall increase of interdisciplinary mobility. In figure 13, on the one hand we plot the ratio of discipline switches over the total number of publications per publication year. Note that the x-axis starts in 2001, as for the first year no disciplines switches can occur. As expected, we can see a steep increase (from 0.2 to 0.37) which does not seem to be flattening out yet. The ratio of broadening events on the other hand, appears to be quite stable over the years (around 0.24).

Figure 13. Ratios of discipline switches and broadenings over the total number of document records per publication year. Note that the x-axis starts in 2001, as for the first year no switching or broadening occurrences can occur.

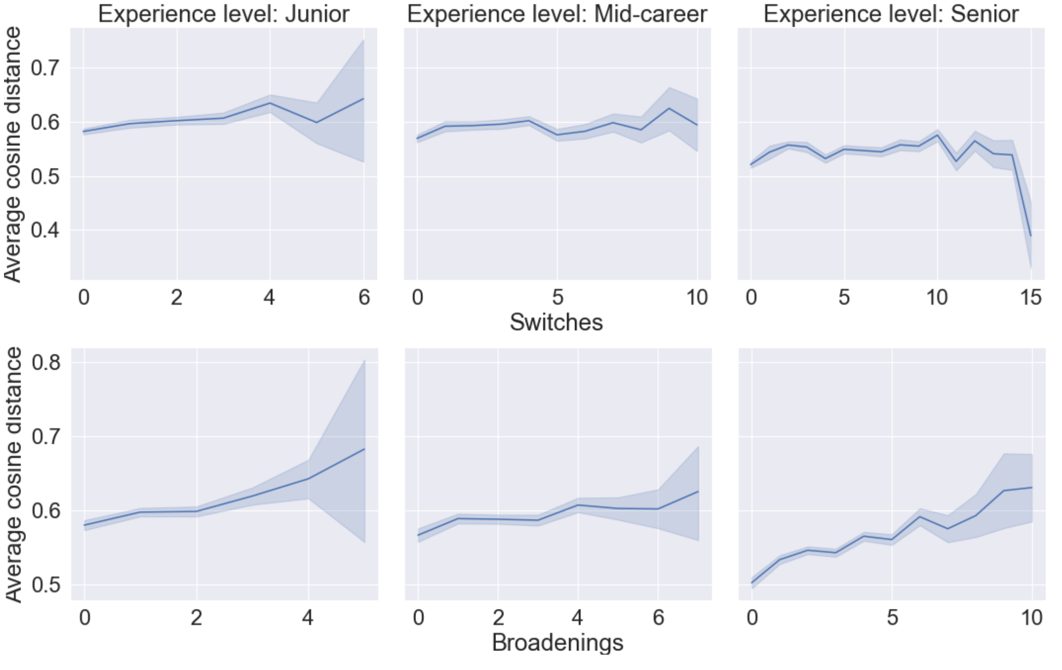


As pointed out in the literature overview, studies on field mobility suggest a relation with the academic age or of authors. Figure 14 displays the relationship between the years of activity of an author and the average cognitive distance she has traveled for each period in which she has been active. The fact that junior researchers are found to be fairly stable in terms of the distance traveled could be related to the high number of PhD researchers included in these groups. The topic of a PhD project is often focused and centered around a well-defined subject of interest. A PhD trajectory typically takes 4 years in Belgium. For the mid-career and senior cohort, we can observe decreasing trends. This can be interpreted as specialization throughout a researcher’s career. It can also be noted that outliers in both directions are found.

Figure 14. Time-series plots with the average cognitive (cosine) distance traveled over periods of activity (one publication year to another). The y-axis represents the average cognitive distance traveled and the x-axis the subsequent periods in which authors have been active. Authors are divided in groups by their number of years of activity. The line indicates the average, and the shaded area the .95 confidence intervals (see Waskom (2021) for details).

The main question of interest here relates to the aspect of disciplinary boundary crossing (switching and broadening) and the cognitive distance traveled. We hypothesize that authors who are more interdisciplinary mobile are also traveling further cognitively speaking. As hypothesized by (Chubin, 1976), for instance, "short distance movement most likely involves specialties ensconced within a single discipline or subfield, while long distance movement may signal movement between specialties located in paradigmatically disparate disciplines" (p. 466). A relation between the cognitive distance traveled and the number of disciplinary switching or broadening occurrences is expected. From figure 15, it becomes clear that the increasing number of discipline switches does not relate a higher cognitive mobility on average. It is only slightly higher for researchers who do switch than for those who don't. For disciplinary broadening, however, this picture is quite different. Those researchers who visit new disciplines throughout their careers, also seem to be more mobility cognitively speaking than those who do not.

Figure 15. Relation between the number of switches (first row) and broadening occurrences (second row) indicated on the x-axis, and the average distance traveled by researchers (y-axis). Researchers are divided in groups based on the number of years of activity registered. Junior researchers 4-7 years of activity, mid-career researchers 8-12 years of activity, and senior researchers 12+ years of activity. The line indicates the average, and the shaded area the .95 confidence intervals with bootstrapping (see Waskom, 2021 for details).



The subset of researchers who have more than 5 discipline broadening counts is fairly small (116). We should also keep in mind that the average distance traveled by the group who broadens often (> 5, average cognitive distance is 0.59) is only slightly higher than that of those who don't (average cognitive distance is 0.55). The high number of broadening events might well be related to cross-disciplinary specialization. For illustration purposes let us take a closer look at the two researchers for whom we counted the most (10) broadening events. One is a professor in agricultural economics and conducts research on agricultural co-operatives, sustainability, and food systems and policy. The other is a full professor in philosophy and publishes on ethical issues related to the natural sciences, evolutionary theory and neurosciences in particular.

7.4 Discussion

Patterns of discipline switching by researchers offer yet another perspective on relations between fields in the SSH. The results obtained by our analyses replicate the common picture of the three cultures (natural sciences, social sciences and humanities), but add nuance in to the assumption that interdisciplinary mobility mainly occurs between fields which are cognitively similar (Hargens, 1986). Accordingly, it is shown that the vast majority of SSH researchers publishes across multiple disciplines. The growth and increasing specialization of the sciences leads to an ever-greater diversity of scholarly communities situated at the intersections of traditional disciplines.

Limitations

While the comprehensiveness of the dataset can be considered a strength for studies of the SSH, the composition of VABB-SHW has its limits. As soon as a researcher is not affiliated to a Flemish SSH research unit, for instance, her writings are not included in the database anymore. Temporal accounts in this case are also limited by the period covered by the database, and not the actual careers of researchers. As a consequence, some noise is introduced in the different career groups. Researchers who have been active in, say 2000-2005 might have well been senior at that time, but ended their careers in 2005. As a consequence, they might have been attributed to the junior group. This is also important to keep in mind when interpreting the results. Career stage information in terms of formal rank (i.e., post-doc, assistant professor, or full professor) would add to our understanding of the findings presented.

Future research

A future direction which we would like to explore is the extent to which researchers are mobile across research specialties (see (Hargens)). Adding a more granular layer based on document clustering for example, cross-specialty publishing of researchers would allow for a more nuanced understanding of how and where cognitive mobility or discipline switching occurs. How, for instance, does the type and newness of subjects in which a researcher is active relate to discipline switching and broadening?

We could for example expect to find that methodologists (e.g. statisticians or data scientists) are more cognitively and disciplinary mobile across the SSH. And does the newness of a field relate to its interdisciplinary identity (see van den Besselaar, 2019; Vugteveen et al., 2014)? "Hot specialties may be most permeable. Many scientists from varying disciplinary backgrounds may gravitate to them without ever being recruited by interpersonal means" (Chubin, 1976, p. 467).

No detailed information on the formal career stage of researchers is available on the author level. It would be genuinely interesting to collect such information, in addition to researchers' (perceptions of their) field membership and motivations for migrating from one field or specialty to another.

7.5 Conclusion

We have looked into disciplinary switching and broadening in the social sciences and humanities by studying publication trends over a relatively long period, and based on comprehensive bibliographic data. The results presented in this study show that quite an extensive network of interdisciplinary traffic of researchers becomes apparent. Indeed, and also according to Jacobs (2013), "disciplines are not silos but rather can be thought of as sharing a dormitory space where they raid each other's closets and borrow each other's clothes." (p. 35). Moreover, in both the discipline switching and similarity network, STEM disciplines were also found to be present. In agreement with what Hargens (1986) found, we could not disentangle a hierarchy of disciplines. The in- and outflow was well balanced for all discipline categories under consideration. While the two networks appeared similar structurally, on a local level we found that authors were not mainly moving between disciplines which were cognitively most similar.

As the results also indicate, over time researchers are switching between discipline categories more often, and this is not related to the actual cognitive distance they travel. The frequency of interdisciplinary mobility does not lead to a higher cognitive mobility. An explanation for this finding is that more sub-disciplinary groups or specialisms are manifesting themselves on the intersections between disciplines. On the contrary, when researchers broaden their disciplinary scope throughout their careers they seem to be traveling further cognitively. The breadth of interdisciplinary mobility appears to be positively related to cognitive mobility. This last finding could be interpreted as a consequence of researchers who are active in a new specialty and by doing so explore the boundaries of multiple fields. While we have been able to identify different groups of cognitively mobile scholars and approached the aspect of cognitive mobility as related to disciplinary or field mobility, many questions remain to be answered.

8 Subject specialties as interdisciplinary trading grounds: The case of the social sciences and humanities

Full reference: Eykens, J., Guns, R., & Vanderstraeten, R. (2022). Subject specialties as interdisciplinary trading grounds: The case of the social sciences and humanities. *Scientometrics, (2022)*. doi: https://doi.org/10.1007/s11192-021-04254-w.

The growth of science is characterized by a broad and increasing variety of both small and large specialties. “As specialization was introduced into scientific work”, the sociologist Emile Durkheim noted already more than a century ago, existing Leonardesque ideals about intellectual labour had to be given up: “science, carved up into a host of detailed studies that have no link with one another, no longer forms a whole” (1893/2014, p. 279). The modern world of science cannot be described as a unitary one; heterogeneity and diversity are the counterpart of increasing specialization (Campbell, 1969; Abbott, 2010).

Disciplinary differentiation, to which Durkheim referred, became widely institutionalized in the 19th century and still plays a major role in the present era. Disciplines, such as physics, biology, history, psychology, or sociology, still give structure to the world of science at both the cognitive and the organizational level. But this disciplinary differentiation is also (and increasingly) called into question. Calls for ‘more integrated’ research, which blur the boundaries of the traditional disciplines, have gained popularity in a variety of policy contexts. A broader interest in interdisciplinary exchanges and cross- or post-disciplinary developments has also emerged (e.g., Graff, 2015; Jacobs, 2013). An important ensuing question for scientometric research is how we can take stock and make sense of the diversity and heterogeneity in the contemporary world of science. Fine-grained analyses of the complex structures of science are also a prerequisite for assessments of ongoing developments.

8.1 Literature review

In scientometrics, diverse studies on the impact and role of interdisciplinarity in science have already seen the light of day. Citation network analysis has repeatedly been used to depict and measure the degree of interdisciplinarity of scientific journals empirically (Karunan, Lathabai, & Prabhakaran, 2017; Leydesdorff, Wagner, & Bornmann, 2018, 2019). Metadata about publications have been used to analyse the historical trajectory of specific disciplinary or interdisciplinary specialties (e.g., Núñez et al., 2019; Porter & Rafols, 2009; Vandermoere & Vanderstraeten, 2012; Vanderstraeten & Vandermoere, 2015; Vugteveen et al., 2014).

Vugteveen et al. (2014) discuss river research as an emerging, potentially interdisciplinary field. The authors show that many fields are involved in the development of its knowledge base; "although river science operates in a traditional disciplinary mode […] various research topics represent a combined contribution of disciplinary research, which implies multi-disciplinary research efforts at the operational level" (Vugteveen et al., 2014, p. 93). Schoepflin and Glänzel (2001) drew a similar picture for the specialty of scientometrics. Different disciplines contribute to this field, but over time the balance appears to be shifting toward case-studies and methodology. McCain (1998) likewise sketched a map of early neural network research as an emerging interdisciplinary field.

The interdisciplinary structure of the social sciences in particular has also been the subject of a range of data-driven empirical studies: Wright (2011), for example, discusses the specialty of public administration as an interdisciplinary field; van Baalen and Karsten (2012) present the evolution of management studies as an interdisciplinary field; Ostrom (2007) sketches the development of institutional analysis as an interdisciplinary field, etc.

Some case-studies indicate that multi- or cross-disciplinary (social sciences and humanities) specialties may over time evolve into fully fledged disciplines with their own journals, conferences, societies and academic curricula (e.g., McCain, 1998; Vandermoere & Vanderstraeten, 2012, Vanderstraeten, 2021). Such a trajectory has been studied bibliometrically by van den Besselaar (2019), who argues that specialties are in constant flux, integrating knowledge from and disseminating it to other fields. While all specialties show overlap with different neighbouring disciplines, Van den Besselaar hypothesizes that they undergo a similar life trajectory, developing from a multi-disciplinary site of knowledge exchange over a certain research question to interdisciplinary knowledge integration and finally to a discipline-like structure with a delineated research program and accompanying communication infrastructures (journals, conferences, curricula, etc.). According to this hypothesis, one should be able to find that different specialties are more or less interdisciplinary at some point in time than others.

A somewhat older study, which is to our knowledge the only larger scale study to date on this topic, has been conducted by Small and Crane (1979). The authors show that specialties in social sciences as well as in STEM fields play an intricate role in forms of knowledge transfer that bridge different disciplines. They use co-citation analysis to identify clusters (specialties) in the natural and social sciences and discuss the interconnectedness of specialties in both broader fields of study. While they hypothesized that specialties in social sciences would be more interconnected, bridging different disciplines, they found that fields in natural science are also highly connected. For social sciences in particular, the authors find that sociology is highly interdisciplinary. Specialties which belong to this discipline have a high tendency to be linked (co-citation-wise) to specialties which belong to other disciplines.

From the perspective of disciplines, Small and Crane (1979) thus argue, the specialties are a connecting component. They bring scholars and research questions from different disciplines together. However, some disciplines can be regarded as being more 'open' or 'closed' to this sharing of subjects. And while sociology is often depicted as an open discipline, economics is often used as an exemplar of a closed, insular social science discipline. Truc et al. (2020) have recently confirmed this perception. The downside of relying on citation data is that one is limited to what is included in citation indexes, thereby marginalizing locally oriented research, non-English publications, and publication types other than journal articles. Another option is therefore to study the disciplinary and interdisciplinary diversity of science on the basis of publication text.

Building upon this literature, we here intend to analyse the disciplinary and interdisciplinary diversity of science. We particularly aim at developing a bottom-up approach to analyse the diversity of interdisciplinary orientations of research specialties in an encompassing multilingual dataset of publications in all the social sciences and humanities. We also hope to provide new input for and advance the long-lasting discussions about the 'best possible' structure of the complex world of science (Campbell, 1969; Abbott, 2010).

Study aims

In this study we investigate if research specialties fulfill a role in bridging different disciplines, both from a social (or organizational) perspective and from a cognitive angle. By investigating the usefulness of the typology of disciplinary vs. multidisciplinary specialties (van den Besselaar, 2019; van den Besselaar & Leydesdorff, 1996) for the specific context of the SSH, we first broaden our understanding of research specialties in this case. Research specialties are operationalized as textually coherent clusters of documents, dealing with a similar subject of interest (e.g., developmental psychology, technometrics, project planning, bilingual education, etc.). Similar to Small and Crane (1979) we also study the openness of different disciplines based on their constituent subject specialties.

The idea of a scientific discipline is an ambiguous one (Bjorn Hammarfelt, 2018; Sugimoto & Weingart, 2015). This has resulted in numerous studies investigating the usefulness of different methods to operationalize the concept (e.g., Guns et al., 2018; Sīle, Guns, Vandermoere, Sivertsen, & Engels, 2021). Organizational and cognitive approaches based on either the departmental affiliation of authors or the cognitive scope of journals and/or individual publications are popular. In this study we use both methods. The interdisciplinarity of subject specialty is operationalized by studying if two or more disciplinary categories are present within a cluster of documents (either in terms of the cognitive scope of the publications, or the affiliation of the researchers who authored a study).

By investigating the alignment between the social (or organizational) aspects of disciplinarity and the cognitive scope of research specialties, we also add a piece to the puzzle which has been posed in earlier work on SSH discipline categories and their alignment (i.e. What are the different factors at play in connecting different disciplinary classification systems?) and which has thus far received little attention in bibliometric research (Guns et al., 2018). We also study how the diversity in terms of organizational affiliations of authors relates to the cognitive diversity in terms of the scope of the publications present in the clusters.

Outline

This research builds further on a conference proceedings paper presented at the ISSI2021 conference (Eykens, Guns, & Engels, 2021). In this study we investigated if word and document embedding techniques (Word2Vec and Doc2Vec) yield qualitative results when used for document clustering in the context of SSH. It was found that these methods yield qualitatively better outcomes in terms of textual coherence than TF-IDF and Latent Semantic Indexing. This study further follows this path by making use of a document vectorization technique from the same family of methods.

In the next section, we first describe the dataset used for our analyses and the steps carried out for cleaning and pre-processing. In the methods section, we describe the document vectorization technique and the dimensionality reduction technique as well as the clustering algorithm which have been applied. The results section discusses the outcomes of the document clustering performed with Top2Vec. We also present some descriptive statistics of the clusters identified (document length, publication type and publication language). The second part of the results section discusses the disciplinary diversity or ‘interdisciplinarity’ of the subject specialties. In the third section the disciplinarity is discussed and the alignment between the organizational and cognitive orientation. Following the results, we present some noteworthy limitations and discuss the implications of our results for research into interdisciplinarity and specialization. The concluding section looks forward and sketches an agenda for further research into the role played by research specialties in terms of bridging the so-called disciplinary silos.

8.2 Dataset: The Flemish Bibliographic Database for the Social Sciences and Humanities (VABB-SHW)

The dataset which is used for the analyses is part of VABB-SHW, the Flemish Bibliographic Database for the Social Sciences and Humanities, hereafter referred to as VABB. VABB covers publication years 2009–2018 and consists of 78,512 metadata records for publications authored by researchers affiliated to an SSH department or faculty at one of the five Flemish universities (for details, see: Engels & Guns, 2018; Verleysen et al., 2014). The most important publication types are included in VABB: peer-reviewed journal articles, book chapters, monographs, edited books, and conference proceedings, in all languages. Currently, two disciplinary classification systems are in place – one that is based on the cognitive scope of the journals in which a publication appeared, a classification of the books, or conference proceedings collections (slightly adjusted OECD FOS classification system, for details see: Guns et al., 2018) and one that is based on the organizational affiliation of the authors.

	journal articles	monographs	edited volume	book chapter	proceedings	Total
English	38,962	214	308	1,658	1,575	**42,717**
Dutch	1,514	52	53	119	12	**1,750**
French	237	13	28	29	11	**318**
Spanish	90	3	7	22	9	**131**
German	45	9	7	19	2	**82**
Italian	14	3	3	3	2	**25**
Total	**40,862**	**294**	**406**	**1,850**	**1,611**	**45,023**

Table 11. Number of records per publication type and most prominent language category in the final dataset used for further analyses.

We first selected all records (n = 46,792) for which: (a) we could retrieve an abstract or summary, (b) both disciplinary classifications were assigned, and (c) the abstract concatenated with the title exceeded the lower limit of 65 words. The set was further restricted to publications in the six most frequently occurring languages, viz. English (42,717 records), Dutch (1,750), French (318), Spanish (131), German (82) and Italian (25), leaving a total of 45,023 records (39 documents were written in other languages). The majority of publications in the dataset are journal articles (40,862 or +90%) followed by book chapters (1,850), conference proceedings (1,611), edited books (406) and monographs (294). All cognitive disciplines (cf. OECD FOS codes) are present in the data, with the most prominent categories being clinical medicine and psychology. The share of clinical medicine publications present in the dataset mainly relates to many collaborations of scholars active in social health sciences.

8.3 Methods

To identify subject specialties in the dataset we make use of unsupervised machine learning (text clustering). In a previous study we have explored different text vectorization methods for clustering of SSH publications (Eykens et al., 2021). Established methods like TF-IDF and LSI were compared with more recent techniques (Word2Vec and Doc2Vec) (Le & Mikolov, 2014; Mikolov, Chen, et al., 2013). The latter showed very promising results in this specific case, leading to more textually coherent and well separated clusters of documents. In this previous study, however, emphasis was placed on comparing different methods for English language document clustering in the context of SSH. Here we further follow this path but extend the scope of the clustering to a multi-language dataset. To this end, the Top2Vec algorithm is used (Angelov, 2020).

8.3.1 Top2Vec

As an unsupervised topic modelling or clustering solution, Top2Vec (Angelov, 2020) has two main benefits: (1) it is entirely unsupervised, no pre-set number of topics or clusters is required, and (2) different state-of-the-art pre-trained word and document embedding models can be used, allowing us to apply the algorithm to a multi-lingual set of documents. The main idea behind Top2Vec is to jointly embed documents and words to find topic vectors. No stop-word list, stemming or lemmatization are needed for this. The topic vectors which are computed are then embedded with the document and word vectors, 'with the distance between them representing the semantic similarity' between documents (Angelov, 2020, p. 1).

Document vectorization and embedding

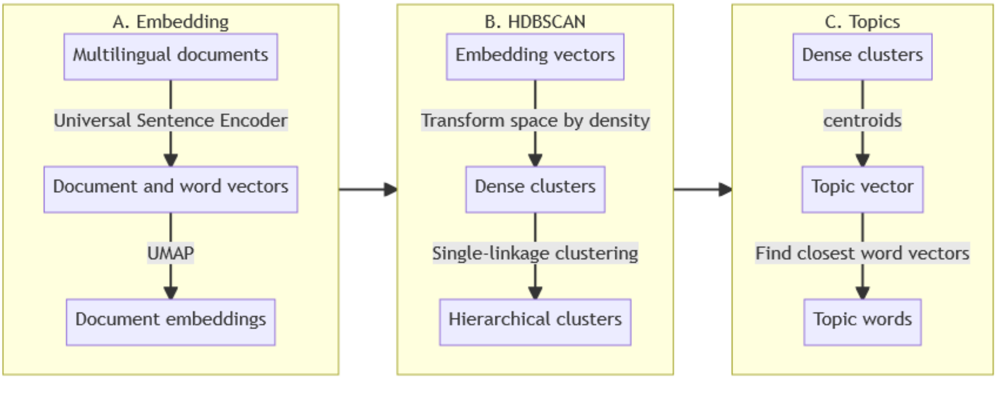
Document vectorization and embedding are represented in box A of Figure 16. A pre-trained text embedding model from the Universal Sentence Encoder family (universal-sentence-encoder-multilingual, obtained from https://tfhub.dev/google/universal-sentence-encoder-multilingual/3) is used for text vectorization (Cer et al., 2018). It has been optimized for cross-lingual retrieval on short text (sentences, paraphrases or short paragraphs) and covers 16 languages. The output is a 512-dimensional document vector for each title-abstract combination. The output of the model, the document and word vectors, can be placed in the same semantic space, where semantically similar documents and words are placed closer together, regardless of language used.

After embedding the documents and words, Uniform Manifold Approximation and Projection or UMAP (McInnes, Healy, & Melville, 2020) is applied for dimensionality reduction. According to Angelov (2020), UMAP serves better in preserving the global structure of the embedding space and scales well to large datasets compared to, e.g., t-SNE. The number of nearest neighbours parameters is set to 15 and the cosine similarity between document vectors is used as a distance metric. The number of dimensions is reduced from 512 to 5. According to the author, this gives the best results for density-based clustering.

Document clustering: HDBSCAN

Hierarchical DBSCAN or HDBSCAN (Campello, Moulavi & Sander, 2013) is applied to the UMAP projection to find dense areas of similar documents (Figure 16, box B). This algorithm is designed to handle noisy data and variable density clusters (Angelov, 2020, p. 7). HDBSCAN is itself a hierarchical extension of the older DBSCAN algorithm (Density-Based Clustering of Applications with Noise) (Ester, Kriegel, Sander & Xu, 1996), which has been used successfully for bibliometric use cases (for a recent example, see: Noichl, 2019). We set the minimum number of documents that should be assigned to a cluster to 15. Finally, the centroids of the clustering outcomes are considered as topic vectors and the closest words as the topic words (Figure 16, box C).

Figure 16*.* Visualization of the Top2Vec (Angelov, 2020) workflow.



8.3.2 Cross-disciplinarity and diversity analyses

To study the disciplinary identity of the subject specialties retrieved by applying Top2Vec, we compute two indicators for the (cross-)disciplinarity and disciplinary diversity (or interdisciplinarity). First, we should briefly highlight the concept of discipline vectors which are used to compute the indicators.

Discipline vectors
Each subject specialty is characterized by the discipline codes that have been assigned to the documents present in a cluster. This is done separately for (I) the organizational classification based on the departmental affiliation of the authors, and (II) the cognitive disciplines. Discipline codes are fully counted, i.e., we do not control for publications with multiple discipline codes. For each subject specialty, this results in two discipline vectors: (I) one with the cognitive counts and (II) one with the organizational counts. Formally, if there are n disciplines, the discipline vector is $X = (x_1, \dots, x_n)$, with element x_i representing the number of publications from discipline i. We will mainly work with the normalized discipline vector $X' = (p_1, \dots, p_n)$, where $p_i = x_i / \sum_{k=1}^{n} x_k$.

(Cross-)disciplinarity and disciplinary diversity
Similar to what is done to calculate the interdisciplinarity of reference lists or collections of journals (Huutoniemi, Klein, Bruun, & Hukkinen, 2010; Leydesdorff & Probst, 2009; Porter & Rafols, 2009), we calculate a diversity index for the discipline codes present in the subject specialties identified with the clustering algorithm. Both for the organizational classification and the cognitive classification. The diversity index takes into account the number of disciplines present within a cluster, as well as the relative abundance of each category.

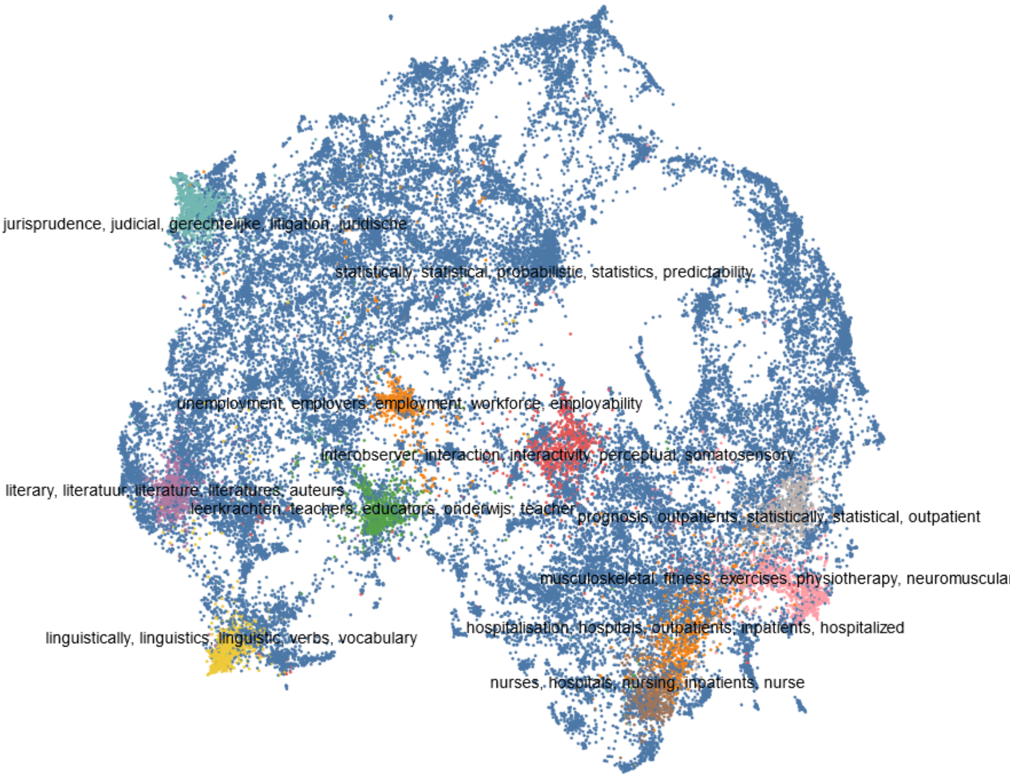
The Hill-type diversity measure $1/\sum p_i^2$ is calculated to get a better understanding of this disciplinary diversity of each subject specialty. It has been shown (Jost, 2009; Zhang, Rousseau, & Glänzel, 2015) that this measure is related to Simpson diversity but better captures our intuition of diversity. Again, Hill diversity is determined for both the cognitive and the organizational disciplines. Since we lack a quantification of how similar these disciplines are, we do not take disparity into account. As the variety and evenness of disciplinary categories present increases, the diversity index also increases.

The (cross-)disciplinarity of a subject specialty from an organizational and cognitive perspective can be approximated using the normalized discipline vector. The largest discipline in the vector is denoted the 'main field' and its p_i value is denoted the 'dominance' of the discipline for the specialty in question. This way of working is similar to what Small and Crane (1979) propose in their work on specialties. To grasp whether a subject specialty should be regarded as either disciplinary or interdisciplinary, we compute the average dominance for all subject specialties. If the dominance of the main field of a cluster is below average, we regard that cluster as interdisciplinary, and vice versa. For each discipline group of specialties, we also calculate the standard deviation from the average.

8.4 Results

In the first part of the results section, we describe the outcomes of the document clustering. In the second part, we present the results regarding the disciplinary diversity of the subject specialties. We present the distribution of the Hill Diversity index for all subject specialties, and show how organizational diversity present in a subject specialty relates to its cognitive diversity. The last part discusses the (cross-)disciplinary identity of the subject specialties. We elaborate further on the question whether some disciplines can be regarded as being more 'open' to share subjects with other disciplines to further contextualize the findings presented in part 2. The alignment between organizational and cognitive disciplines in subject specialties is discussed as well to shed light on discrepancies between these two systems.

Figure 17. UMAP projection of the document vectors. The 11 largest clusters are highlighted in color and labelled with the first five topic words (according to Top2Vec).



8.4.1 Clustering outcomes

Figure 17 shows a 2D UMAP projection of the clusters. The largest clusters are labelled with the most prominent keywords. Note that these are quite general. This mainly has to do with the size of these larger and more generic clusters. 246 subject specialties are identified with Top2Vec, with an average size of 183 records (median 127.5). The vast majority of clusters consists of less than 221.5 records (75th percentile). The smallest cluster contains 28 records (subject specialty # 245, deals with mathematics education and didactics), while the two largest clusters contain over 1,000 + publications. The distribution over cluster sizes is heavily skewed, with a majority of clusters ranging from 28 to 221.5 records.

8.4.2 The disciplinary diversity of subject specialties

From figure 19 it becomes clear that the diversity found for the subject specialties varies a lot. For both classification systems the centre of the distributions lies around 4. For simplicity, we will treat 4 as the threshold value for low/high diversity. Both distributions are skewed to the right, meaning that the majority of specialties is more diverse than the mean of the distribution. The variance is somewhat larger for the cognitive classification system due to the higher number of disciplinary categories present in this system.

We now turn to the question how disciplinary diversity on the input side is related to disciplinary diversity on the output side. Figure 20 displays this relation in a scatterplot. We see that there is a positive linear relationship between the two. The assumption of homoscedasticity however, is not met. For the specialties with a relatively low diversity (organizational and cognitive below 4), the dots are still quite close to each other, but going further up on both axes, the variance quickly increases. This hinders an interpretation in terms of linear regression with confidence intervals. For further qualitative interpretation, we have divided the plot into four parts.

Disciplined specialties (72 clusters)
Part A, x-axis below 4, and y-axis below 4: low cognitive diversity, low organizational. This part of the plot covers the leftmost half of the distributions presented in figure 4. Scholars contributing to these subject specialties are mostly located within the same disciplinary branch of a university, and the output they produce is quite homogeneous cognitively speaking; they can be termed disciplined specialties. The idea here aligns with what we commonly think of when we use the term

Figure 18. Distribution of clusters over different cluster sizes.

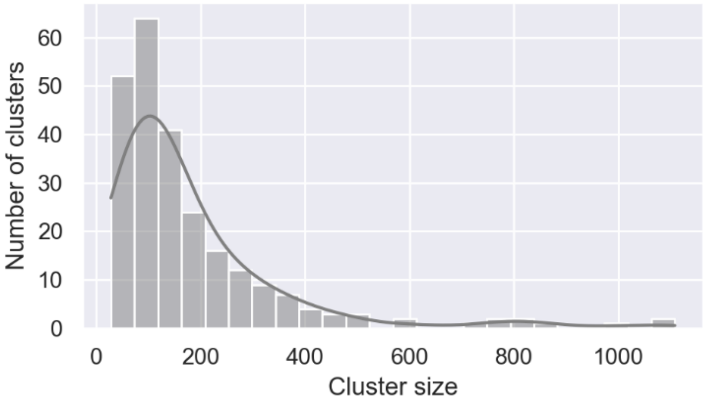
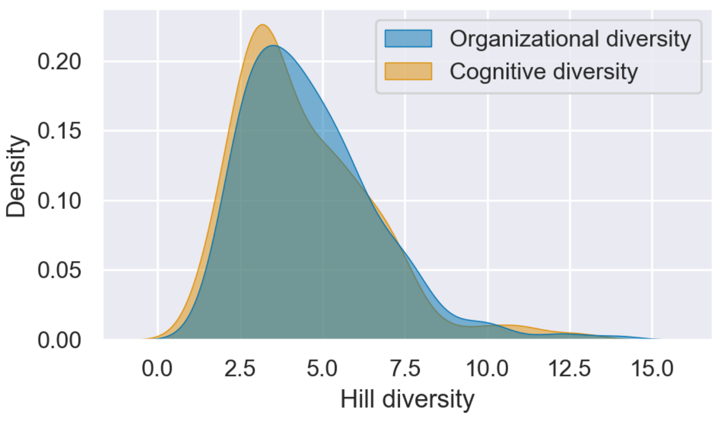


Figure 19. Density distribution of disciplinary diversity for all subject specialties. Based on the organizational affiliation of authors and the cognitive disciplinary classification.



discipline – or when we apply traditional disciplinary classification systems. People working within a similar discipline department-wise publish content which is similar to that of their colleagues working at the same department cognitively speaking (in terms of methods applied, concepts and theories used, and subject matter studied). Subject specialty #177 is one such example. This disciplined specialty deals with research on the life and work of Saint Augustine. Most authors contributing here work at a theology department and publish in theology or philosophy journals.

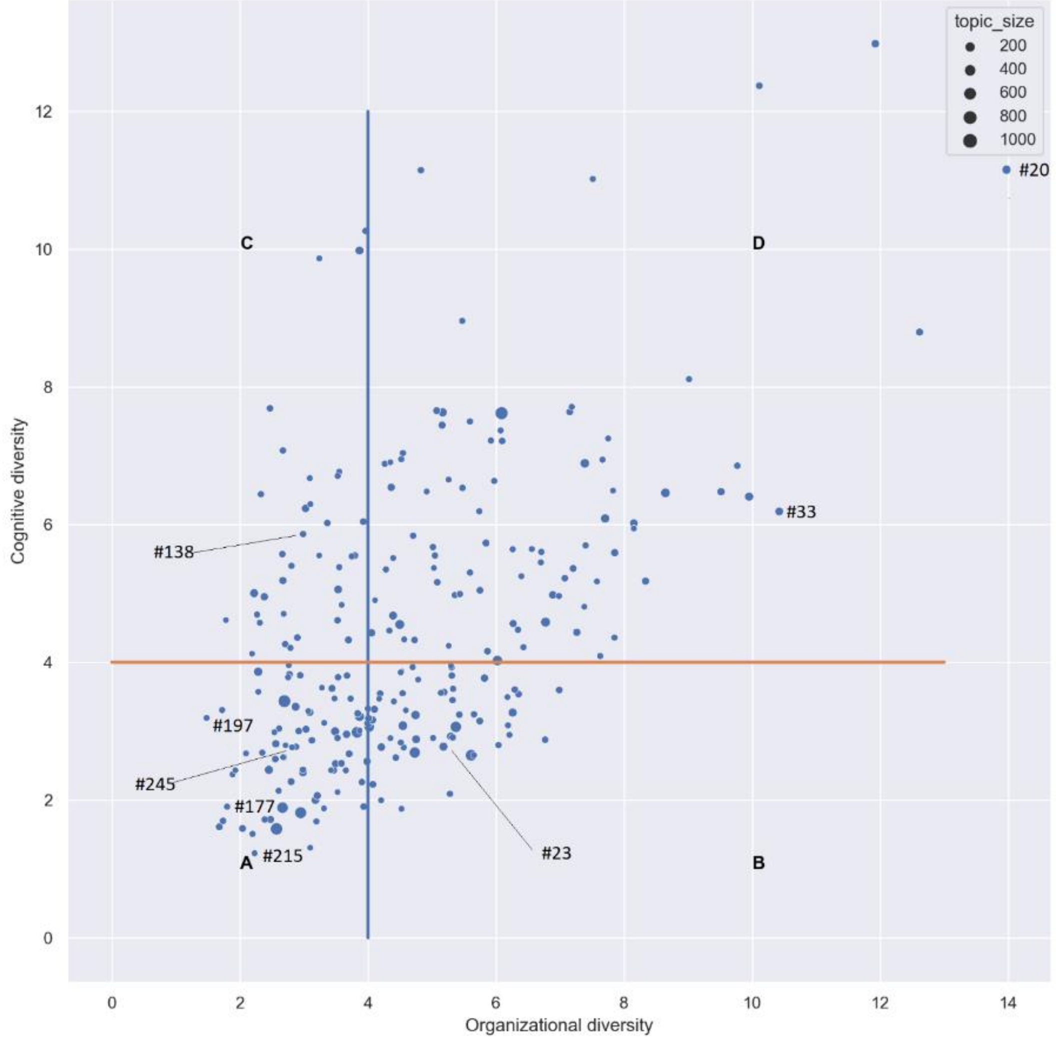
Fractured disciplinary specialties (55 clusters)

Part B, x-axis 4+, y-axis below 4: low cognitive diversity, high organizational diversity. The cognitive diversity is below the median, but the organizational diversity is high. Fractured disciplinary specialties are populated by a group of specialists coming from different disciplinary working grounds, but contributing to the knowledge base of a common traditional discipline. This might be the case for multi- or interdisciplinary specialties which have been growing towards a discipline-like structure, with devoted journals, but have not yet been further institutionalized (i.e., departments and devoted research institutions). Larger multi- or interdisciplinary research projects might be captured by this category. These specialties have the potential to become disciplined institutionally speaking. This is a dense and large group when compared to part C and D of the plot. Cluster #23 is one example. Research conducted within this specialty deals with developmental psychology and the family. More specifically, studies deal with parenting traits and how they influence the psychology and development of the child. While most research gets published in psychology journals (60%), we see authors contributing from medical departments as well as educational sciences, communication studies, and sociology.

Institutionalized cross-disciplinary specialties (35 clusters)

Part C, x-axis below 4, y-axis 4+: a high cognitive diversity, low organizational diversity. While knowledge within these subject specialties is attributed to different disciplinary categories cognitively speaking, researchers populating them mainly come from a similar discipline department-wise. So although the specialty is institutionalized, the knowledge created is cross-disciplinary in nature. These might be inter- or cross-disciplinary research projects conducted within or by a research group working in a relatively novel cross-disciplinary research area. This part of the plot is very scattered and considerably smaller in terms of the number of clusters present here. Cluster #138 deals with statistical techniques and technometrics. The largest share of researchers in this cluster work at a business and economics department, but the research they produce is being published in mathematics journals, psychology, engineering, and business and economics. The methodological nature of the research conducted here makes it relevant for many fields.

Figure 20. Scatter plot showing the relationship between the organizational disciplinary (Hill) diversity (x-axis) and the cognitive disciplinary diversity (y-axis) of subject specialties. Dot size is equivalent to the size of the subject specialties represented.



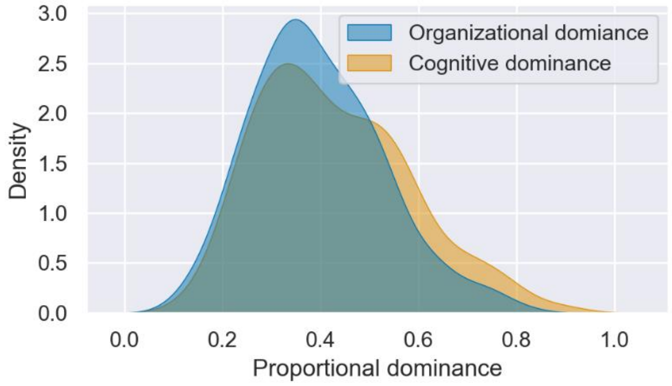
Fractured cross-disciplinary specialties (84 clusters)
Part D, x-axis above 4, y-axis above 4: high cognitive diversity, high organizational diversity. This is clearly the largest part of the plot, and the least dense. Scholars are working at different departments and publish in many different disciplines cognitively speaking. One

extreme case is cluster #20. This subject specialty deals with cultural history. The two main organizational categories with which the research gets classified are tellingly, social sciences general and humanities general. Research gets published in journals from sociology, education, literature, psychology, history, etc. Publication types and languages used are diverse as well. Subject specialty #33 is another example of this type. This cluster is concerned with ethnic and migration studies. Again, not only the disciplinary categories, but also the publication types and languages used to communicate the findings are remarkably diverse.

8.4.3 The cross-disciplinarity of subject specialties

As indicated by the short literature review and the findings presented above, many subject specialties are fulfilling an important function in bringing together knowledge from different disciplines. But to what degree are specialties in the social sciences and humanities disciplined by specific categories? And are some specialties more 'disciplined' than others? We also pose this question the other way around: Are some disciplines more 'protective' of their subjects? Similar to the method used by Small and Crane (1979), the latter question is explored by dividing the set of subject specialties into disciplinary categories based on the disciplinary dominance (highest proportion of disciplinary category present in a cluster). Let us first look at the degree of cross-disciplinarity across all clusters in Figure 21.

Figure 21. Density distribution of disciplinary dominance for all subject specialties. X-axis indicates the proportional dominance. Based on the organizational affiliation of authors and the cognitive disciplinary classification.



While in both cases the average disciplinary dominance lies somewhere around .4 for the majority of clusters (meaning that the main disciplinary category accounts for only 40% of publications), we find that for both the organizational and the cognitive disciplinary categories, no single cluster is entirely dominated by one disciplinary group. The main conclusion which can be reached from this picture is that all subject specialties found in our dataset are inherently cross-disciplinary, with some outliers in both directions.

For the organizational classification, subject specialty #197 (also labeled in figure 20) sticks out with a dominance value of 0.81 for business and economics. This specialty deals with resource-constrained project planning and scheduling. The cognitive dominance, however, lies around the average value of 0.4. So, although this subject specialty appears to be highly disciplined from an organizational perspective, cognitively speaking, contributions are made to engineering fields, computer and information sciences, mathematical modelling, media and communication studies, etc. A similar case can be made for a highly disciplined subject specialty in terms of the cognitive disciplines it relates to.

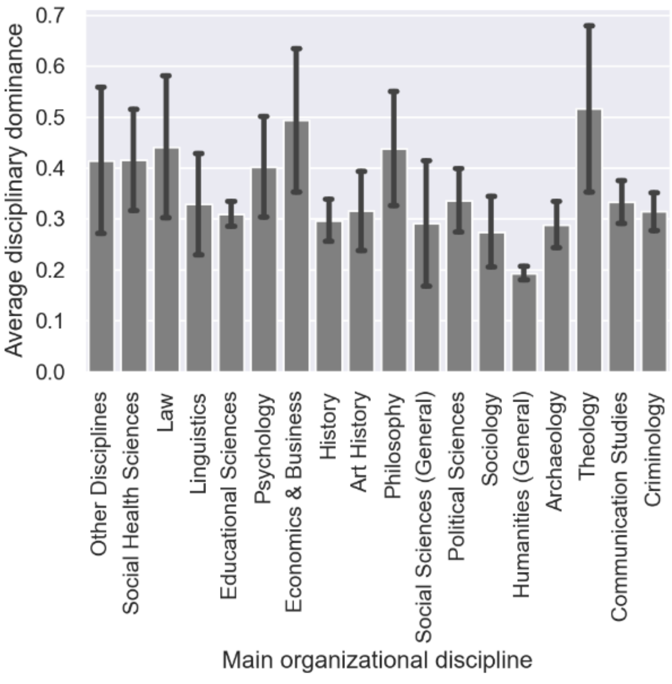
Subject specialty #215 has a disciplinary dominance of 0.9 cognitively speaking, producing work mainly within the scope of

Earth and related environmental sciences. Organizationally, however, researchers affiliated to archaeology, history, and business and economics departments contribute here. While cross-disciplinarity on the input side and on the output side are clearly correlated (cf. Figure 20 and 21), these examples illustrate that they are not interchangeable and tell a different story about either the specialization or (cross-)disciplinarity of a subject specialty.

8.4.4 The openness of disciplines

We wanted to explore to what extent disciplines are 'protective' or 'dominant' over certain subject specialties. To study this question, we first assigned each subject specialty to its main field (the dominant disciplinary category). This has been done for both the organizational classification and the cognitive classification. We first look at the average disciplinary dominance in terms of their proportion. Next, we take a look at how many subject specialties for a particular disciplinary category have been classified as being multidisciplinary. A subject specialty is considered multidisciplinary if the average disciplinary dominance of the main discipline is below the average dominance calculated for all subject specialties (threshold is set to 0.4 for both organizational and cognitive classifications).

Figure 22. Organizational disciplinary dominance over subject specialties. Subject specialties are divided over disciplinary categories according to their main disciplinary category proportionally speaking. Next, the average proportional dominance of the specialties is calculated for that disciplinary category. Error bars are included which show the standard deviation.



The subject specialties in which researchers are affiliated to a theology research group or department are the most 'disciplined'. Apart from philosophy this is the only traditional humanities discipline for which this the case. Linguistics, art history and archaeology all seem to be present in subject specialties which are extensively shared with researchers from other disciplines, with at the lowest end of the spectrum archaeology. This mainly has to do with the high number of researchers from STEM fields as well as history, who are also interested in either the objects studied or discovered, or the technology involved in archaeological research.

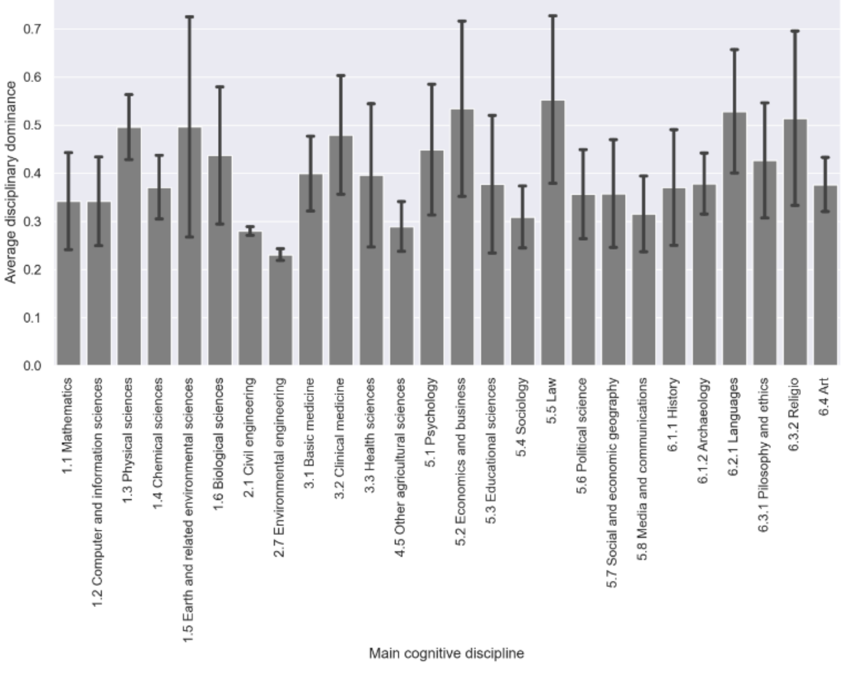
In the case of the social sciences, we see that sociology is the most 'open'. Subjects for which sociologists are the most dominant group of researchers seem to be populated by quite a few scholars from other disciplines as well. Professional fields like law and social health sciences on the other hand appear to be quite disciplined in terms of the subjects which they study. For the social sciences, economics and business is the most disciplined. This last finding is in line with what previous studies have shown about the 'insularity' of economics and business scholarship (Truc et al., 2020).

In figure 23 the results are displayed for the cognitive scope of the publications (or in case of journal articles of the journals in which a publication has appeared). A slightly different picture emerges for the humanities, with languages and literature having the most disciplined subjects. In line with the organizational system, religion again scores quite high. Art, history and archaeology seem to be the least disciplined. For the social sciences, sociology appears the least disciplined. Economics and business and law again score fairly high (see also: Vanderstraeten & Vandermoere, 2021).

8.4.5 The alignment of disciplinary classifications for subject specialties

An important question about research output in general and interdisciplinarity in particular relates to the extent to which disciplinary classifications are correct or relevant. Although we have already shown that disciplinary categories do not align well with subject specialties, it remains to be seen if the dominant disciplinary categories present in a subject specialty relate to each other. For example: are the studies conducted in a subject specialty which is predominantly populated by, say, psychologists also mainly published in psychology journals? We try to answer this question by providing an alluvial plot in figure 24. For the subject specialties which are mainly populated by researchers from the humanities (i.e. history, theology, archaeology, and linguistics in particular) we can see that this alignment is nearly perfect, with art history and philosophy as exceptions. For art history we see veins from the organizational side of the chart (left) to civil engineering on the right-hand side and social and economic geography. Philosophy is connected with biological sciences and medical sciences.

Figure 23. Cognitive disciplinary dominance within subject specialties. Subject specialties are divided over disciplinary categories according to their main disciplinary category proportionally speaking. Next, the average proportional dominance of the specialties is calculated for that disciplinary category. Error bars are included which show the standard deviation.



Whereas it was indicated above that economics and business is less cross-disciplinary than their neighbouring disciplines in social sciences, the alluvial chart displays clear links between economics and business and STEM fields as well as social and economic geography. This finding is in line with previous research stating that economics and business is a knowledge exporter (Yan, Ding, Cronin, & Leydesdorff, 2013). Psychology is linked to sociology and educational sciences as well as medical fields. Scholars from communication studies – arguably one of the youngest fields in social sciences – contribute to subject specialties from computer and information sciences (e.g., scientometrics), psychology, economics and business, and to communication studies. While the alignment between the disciplinary categories of the subject specialties is in most cases straightforward, we do find some interesting and noteworthy cross-overs. The inter- or cross-disciplinary subject specialties can be identified by looking at these discrepancies.

8.5 Discussion

The results presented in this study show that all identified subject specialties are interdisciplinary to some extent, with important variations between them. By contrasting two different disciplinary classification schemes, additional insights have been gained about the different ways these two systems co-exist within subject specialties in the social sciences and humanities in particular. We have (briefly) presented a simple typology that could serve as a guideline for research evaluation exercises in this light. Disciplined specialties on the one hand seem to prove the case in point for disciplinary evaluations that work. The alluvial chart, however, shows that some disciplined specialties, when regarded in terms of author affiliation, do produce knowledge outside of the organizational discipline. Which disciplinary evaluation system should be applied in such cases? What about institutionalized cross-disciplinary specialties on the other hand, which are being evaluated based on author affiliation and collaborations between authors with different disciplinary backgrounds? Is this type of research to be evaluated as being interdisciplinary? Or should it be evaluated based on cognitive diversity of output? Many other questions can be raised with regard to suitable evaluation procedures for the different types of subject specialties and their (inter-)disciplinary identity, but we leave this up to the reader and future work.

Previous research into the cognitive and communicative characteristics of disciplines highlights that scholars' disciplines in some are more open to share subjects or communicate about them with other disciplines. This study partially confirms these results. Subject specialties which are mainly characterized by publications in economics and business journals – or which are mainly populated by scholars affiliated to a business and education department – show less overlap with other disciplines (Truc et al., 2020), while sociology has often been quoted as a very 'open' discipline, sharing many subjects with other disciplines (Abbott, 2001; Small & Crane, 1979). Guns et al. (2018) have studied the alignment between the cognitive and organizational disciplinary classification systems. In this study we have tried to expand on these findings by adding subject specialties to the story. We show that specialties indeed play a peculiar role in connecting different categories, sometimes leading to noticeable crossovers as well. To gain understanding, however, more in-depth qualitative research would be appropriate.

In sum it is shown that science is evolving into a (hyper-)specialized reality in which broad discipline categories quickly become irrelevant when it comes to describing the actual 'on the ground' practices of doing research. Subject specialties that are closely related to a specific discipline might be two entirely different worlds when it comes to, for example, the methods and technical equipment used. As a result, within one single research specialty, scholars originating from different disciplines might be active. We can thus pose the question if a classification according to broad disciplines is still relevant, and if opposing disciplinarity to interdisciplinarity is needed. The results presented in this study suggest that no single subject specialty can be regarded as being purely disciplinary, and that the degree to which it is inter- or cross-disciplinary strongly depends on the classification system or its granularity.

Limitations

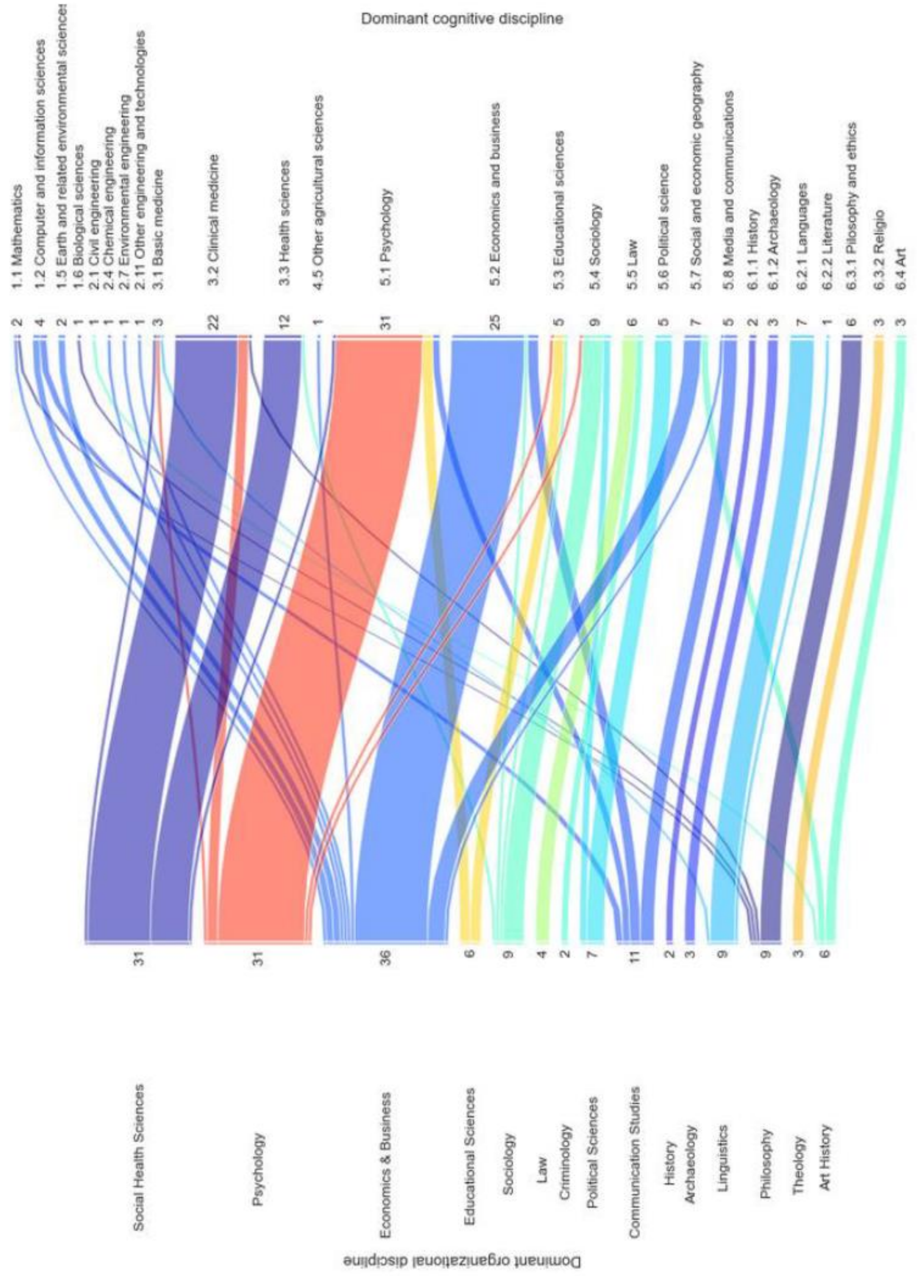
The Top2Vec algorithm used to delineate subject specialties in our corpus has its limits. Being entirely based on textual data, we do not include any information on the social or communicative components of scientific communities (i.e., membership of professional organizations, the same research groups, or citation links). Being able to include more detailed affiliation or collaboration data might be a way forward to further optimize the clustering outcomes. Apart from using titles and abstracts, including keywords would be an interesting sur plus as well.

Universal Sentence Encoders which are applied in this study have not been trained on a corpus which exclusively consists of scientific documents. A consequence of this might be that we find some less well defined clusters. The two largest clusters identified in our topic are points in case. After qualitative inspection it was found that they are rather ambiguous in content. We decided to include these clusters for analyses, as they are the factual outcomes of the clustering algorithm. Additionally, unlike the workflow presented in (Eykens, Guns and Engels, 2021) in the current study we have retained from calculating textual coherence scores for the clustering produced by the Top2Vec algorithm. Because we are working with a multilingual corpus, for some specialties

identified the documents are in different languages. Using a bag of words framework to calculate coherence scores in such cases will yield inaccurate results. A solution might be to use an algorithm to translate all text to English, but this would introduce yet another layer of complexity, potentially introducing translation error.

The use of a local database imposes some limitations as well. Although we are able to sketch a quite comprehensive picture of Flemish SSH scholarship, research is increasingly carried out in an international context. For subject specialties which are mainly oriented at an international level, these data should be complemented with documents from other sources to gain a more appropriate and fitting picture of the entire system.

Figure 24. Alluvial chart displaying the alignment between organizational dominance and main cognitive scope for all subject specialties with authors from social sciences and humanities. The thickness of the veins corresponds to the size of the group. The total number of subject specialties per disciplinary category is indicated next to the labels.



Future research

By plotting the relationship between the two disciplinary classifications systems and their diversity we have only scratched the surface in terms of possible typologies of subject specialties. By using the keywords of the clusters, we could identify, for example, clusters which are devoted to topics related to the Sustainable Development Goals, to specific quantitative or qualitative research methods, etc. and study their disciplinary prevalence and/or cross-disciplinarity.

Fractured cross-disciplinary specialties, which showed a high diversity for both disciplinary classification systems, appeared to exhibit diverse publication practices as well. This could be due to the applied and problem-oriented nature of interdisciplinary research, as is often stated in policy documents and by advocates of interdisciplinary science. It would be worthwhile to further investigate this question. If interdisciplinary research is indeed more problem-oriented, publication patterns might differ for these subject specialties. For the SSH in Belgium, this could for example mean that many findings are communicated in Dutch, French, or German non-peer reviewed outlets complementary to English-language journal articles indexed in commercial citation databases.

Another interesting aspect which could be studied is the inter-specialty or disciplinary mobility of authors. Do closer subjects and disciplines, cognitively speaking, experience or facilitate higher mobility than distant subjects and disciplines? Are there any noticeable differences between where authors start, i.e., in terms of root field or discipline? Are, for example, authors from subject specialties devoted to research methodology more mobile across subjects and/or disciplines? Are they more 'cosmopolite'? Do some disciplines mainly act as knowledge exporters of researchers and others as absorbers?

8.6 Conclusion

Numerous bibliometric studies point toward the intricate web woven between scientific domains or traditional disciplines by interdisciplinary research. Case studies show that specialties play an important role in this regard; they can be denoted as connecting components. To date however, only a few studies have investigated this in a systematic manner. While anecdotal case studies are of utmost importance to uncover complexities and dynamics present within particular fields, the question remained whether and to what extent subject specialties are per definition cross- or interdisciplinary. Especially for the SSH, such evidence was largely lacking. By studying this question, we have shown the importance of subject specialties in this regard; they can be thought of as interdisciplinary trading grounds for traditional disciplines. Subjects and the research questions surrounding them are shared by two or (often) more traditional disciplines.

While different types of subject specialties have been presented here, we found that all clusters play a cross-disciplinary role in one or the other way, with important variations between them. These different types of interdisciplinary or more disciplinary research demand for unique and nuanced evaluation procedures. We have already touched upon the issue of diversity of publication practices of the cross-disciplinary and fractured specialties. Scholars working in subject specialties that publish in different disciplines are exposed to evaluative standards of different disciplines as well. We would also like to stress that no type of subject specialty should be considered 'more qualitative' in terms of the output produced, or producing 'better research' in similar vein. High impact and revolutionizing knowledge has been produced by both disciplined specialties as well as cross-disciplinary specialties (Jacobs, 2013). Both have their merits and should thus be valued as such.

9 Discussion

This thesis had two main objectives: (1) to develop a better understanding of disciplinary differentiation and interdisciplinarity in the social sciences and humanities (primary objective: *parts 1* and *3*) and (2) assess the applicability of text-based classification and clustering approaches to group scientific publications from the SSH on the level of subject specialties (secondary objective, part 2). After summarizing the main findings for each of the objectives, I will discuss the implications for scientometric research into interdisciplinarity and research evaluation. In the fourth subsection, I note the main limitations of the framework and methods used for our analyses. The fifth subsection outlines an agenda for future research.

9.1 Summary of findings

In part 1 of this thesis (chapters *2. Disciplines* and *3. Assessing Interdisciplinary Research in the Social Sciences*), I have first set out a historical and conceptual description of disciplines and their changing role in contemporary science. Specifically, for the SSH, in the second text I have argued that inter- and intra-disciplinary diversity of the socio-institutional and cognitive characteristics of specialties make it difficult not only to indicate what interdisciplinarity entails, but also to transfer research assessment practices from one disciplinary or sub-disciplinary context to another.

Next, to develop techniques for approaching specialties in the SSH, an assessment of different text-based methods has been conducted (see part 2 *Text-based approaches to science classification and clustering*). I have explored the outcomes of supervised machine learning algorithms in terms of accuracy for reproducing fine-grained classifications of documents. While chapter 4 focuses on single-label classification, chapter 5 expands the scope to a multi-label setting. In chapter 6, the last study under part 2, I have experimented with different document vectorization techniques for text clustering. The results of both the supervised and unsupervised approaches are promising, yet, as I will discuss further on, the level of granularity is an important factor to consider (see Table 4 in particular).

Part 3 of this thesis presents two empirical studies. The first one – chapter 7 – discusses the concept of cognitive traveling (e.g., the cognitive distance covered by authors) and disciplinary boundary crossing. As a consequence of disciplinary differentiation, I suggest that research specialties play an increasingly critical role in terms of connecting different disciplinary categories. In line with what is suggested in part 1 of the thesis, I argue that the disciplinary identity of subject specialties can yield fresh insight into the dynamics of IDR. Finally, chapter 8 offers such a systematic inquiry into the disciplinary diversity of subject specialties in the SSH. I conclude that these specialist communities play an important role in connecting the disciplines and all subject specialties involve multiple disciplines to some extent. Variation is found both in terms of the disciplinary diversity of specialist communities as well as the degree to which disciplines are open to share subjects with one another.

Part 1 – Disciplines and interdisciplinarity

RQ1.1*: What are the main functions of disciplines in the contemporary academic system?* Disciplines have long served as the main referents for knowledge production. In the beginning of the 19th century, when discipline-based departmental structures were first installed at universities across Europe (and later in the US), tightly knit research communities operated around a common disciplinary denominator. At that time the discipline, both institutionally and cognitively, could be thought of as a rather coherent institution. The disciplinary communities were relatively small and, in general, shared a common research agenda. This is what Ash calls 'Phase 1 disciplinarity and specialization in the Establishment of the Research University' (2019, p. 624). Soon after their installment at universities however (and after WWII in particular), together with the emergence of national research funding bodies, the academic system grew exponentially (Bornmann, Haunschild, & Mutz, 2021). A consequence of this growth was differentiation, bringing about disciplinary fragmentation into subdisciplines and research specialties. The disciplines started to serve as 'the teaching domain of science, while smaller intellectual units (nestled within and between the disciplines) comprise the research domain. Within the sociology of science, these units have been termed "scientific specialties".' (Chubin, 1976, p. 448).

While new cognitive or specialist communities were formed around subject (e.g. social movement studies, gender studies, peace studies, etc), research methodology (e.g. social network analysis, ethnography, etc.), theoretical issues, and epistemological debates (e.g. structuralist vs. constructivist), the organizational system which up until then nearly paralleled the cognitive communities changed its function. From harnessing disciplinary research agendas together with stipulating disciplinary curricula, its main function has now moved to organization education and bringing together problem-oriented research groups or centers.

Due to internal changes and external influence, the disciplinary differentiation brings about an ever-rising number of sub-disciplinary and interdisciplinary specializations. These knowledge communities develop their own specific knowledge creation and production practices leading to intra- and interdisciplinarity diversity in terms of esteemed publication formats, languages, epistemological perspectives, etc. This loss of internal unity in science makes the distinction between what is to be considered disciplinary or interdisciplinary research more difficult. "Disciplines are not only the product of internal practice, but also their knowledge relations to and differentiation from other disciplines. Intra disciplinary fragmentation or differentiation cannot be detached from a discipline's relation to other disciplines. The differentiated knowledge relationships held by distinct intradisciplinary fragments to other disciplines" (Aris, 2021, p. 175).

<u>RQ1.2</u>*: How are disciplines approached in bibliometric research?* I have argued that disciplines are not coherently structured around a common research program as they were in the disciplinary phase. Differentiation and specialization make disciplines cognitively heterogeneous entities and 'moving targets' (i.e. new disciplines emerge, some decline or disappear, some merge).

Yet, bibliometric studies suggesting indicators of interdisciplinarity regard disciplinary categories as homogeneous and clearly demarcated, stable units (for an answer to <u>RQ1.3</u>*: Which indicators* *and procedures exist for approaching and assessing interdisciplinarity?*) The idea that disciplines are cognitively homogeneous structures which are well aligned with institutional forms strongly resonates in this approach.

RQ1.4*: Should alternatives be considered for evaluating interdisciplinarity?* I argue that we should take into consideration the resulting intra- and interdisciplinary diversity of social and cognitive characteristics. These are not well represented by fixed discipline-based ontologies used in purely quantitative assessments. Variations between disciplines and subdisciplines should be taken into account when evaluating research. Additionally, it is argued in chapter 3 that, in the case of interdisciplinary research, a dialogue is needed between those who conduct interdisciplinary research and those who evaluate it. Appropriate evaluation of interdisciplinarity is not given but made. To inform a co-creation model of evaluation procedures, the seven principles which have been proposed by Julie Klein (2008) have been reiterated in this chapter.

Part 2 – Text-based approaches to science classification and clustering

Sub-disciplinary communities and specialty groups, often in close interaction with communities outside of the academic system, are becoming the primary referents when it comes to the (co-)creation of research. These sub-disciplinary or specialist communities share institutional characteristics with the disciplinary communities from the 19th century (i.e., a devoted research program, specialist and dedicated journals, shared conceptual frameworks and theories, devoted conferences), but have become detached from their organizational roots. The departments today are the residency of sub-disciplinary and specialty groups, and serve as a referral address instead of the primary referent for a researchers' cognitive identity.

In Flanders, a more fine-grained disciplinary classification system has been introduced in order to be able to take stock of these more granular communities, the VODS or Flemish Research Discipline Standard (Vlaamse Onderzoeksdiscipline Standaard in Dutch). The classification system is based on the OECD Fields of Science discipline code list, but adds two more granular levels. A database wide inclusion of these additional categories, both on the level of sub-disciplines and specialties, would allow us to study disciplinary dynamics and research practices for the communities associated to them. As we have learned, a top-down implementation of this classification system is not straightforward.

For the social sciences and humanities, it is known that publication and referencing practices differ from STEM fields, making citation-based classification or clustering approaches less suitable for application to these fields. Therefore, a second topic of this thesis related to improving our understanding of what the use of automated text classification and clustering could be for approaching granular cognitive or specialist communities in the SSH.

RQ2.1: *To what extent can we make use of supervised machine learning to reconstruct fine-grained, specialty level classifications of social science publications*? The first two studies in part 2 of this thesis discuss the results of supervised text classification algorithms when applied for fine-grained document classification based on the VODS (i.e., on the specialty and subdiscipline level for prominent social science disciplines). Whereas previous studies have shown that such supervised classification approaches yield satisfactory results when applied on broad disciplinary classifications, work studying fine-grained classifications for scientific text has been lagging behind, especially for social sciences and humanities (exceptions exist in the context of medical and biomedical sciences).

On the level of research specialties (in sociology) we have studied four document classification algorithms in a single label assignment context: Multinomial Naïve Bayes, Support Vector Machines, Random Forests, and Gradient Boosting, in addition to a simple TF-IDF vectorization method. We found that Gradient Boosting yields a high accuracy score (.80) and decided to move forward and investigate the potential for multi-label classification. The sociology dataset was expanded by adding sub-disciplines and specialties from educational sciences and business and economics. Domain experts were consulted to validate the classification of the titles and abstracts. Random samples of documents from each discipline were reclassified by the experts and it was found that a relatively good consistency was achieved on the sub-disciplinary level, comparable to earlier evaluations of multi-label classification by professional indexers or field experts in a similar context. The results on the most granular level were unsatisfactory and therefore the machine learning algorithms were evaluated on the sub-disciplinary level.

Gradient Boosting and Multinomial Naïve Bayes were then used to classify the expanded dataset. To achieve a multi-label classification, the algorithms were 'chained'. A Classifier Chain model takes into account label dependency. For each disciplinary dataset, different scoring metrics were calculated to assess the precision of the two algorithms. We found that Gradient Boosting performs well, with F1 scores similar to those found for the experts. While promising in itself, the Gradient Boosting chaining approach was not applicable on the most fine-grained level. Since Gradient Boosting proved quite time-intensive and not entirely suitable to group documents on the most granular level, we also wanted to explore unsupervised document clustering. These are machine learning approaches which do not rely on predefined classes or categories, but build document groups from the bottom up.

<u>RQ2.2</u>*: What are efficient ways of text clustering to construct bottom-up classifications of social sciences and humanities publications?* Unsupervised document clustering without making use of a predefined list of categories is yet another way of approaching the operationalization of fine-grained cognitive communities or subject specialties. In the third study under part 2 we have explored the suitability of this method for clustering English language SSH documents from VABB-SHW. We applied k-means clustering – a straightforward algorithm commonly used for unsupervised learning – to group titles and abstracts of the publications. The main aim here was to assess the textual coherence of the clusters for established and newer document vectorization techniques. It was found that, when compared to TF-IDF and Latent Semantic Indexing (LSI), Word2Vec and Doc2Vec performed slightly better on average.

These latter two vectorization methods are part of a newer generation of word and document embedding techniques that use neural network models to associate words, yielding the promise of obtaining higher accuracy. The stream of machine learning research which is developing around these neural network-based document vectorization approaches has resulted in techniques which are capable of embedding words or concepts, paragraphs, and documents from multiple languages into the same vector space. Recently, different (multilingual) topic modeling approaches have been developed which make use of clustering and neural network-based models for document vectorization showing promising results (Top2Vec, LEGAL-BERT, BERTopic, etc.), some of which are specifically designed for or trained on text from particular subfields in science. In a context of multilingual scholarship, as is the case for the SSH, we can expect that these methods will attract a lot of attention from the scientometric community.

Part 3 – The interplay between – and isolation of disciplines in the social sciences and humanities

As argued in the introductory chapters, it is expected that broad discipline categories capture the everyday practices of knowledge production only to a limited extent. They do however still function as overarching and abstract reference points. Discipline-based departments are, for example, still the default at most universities, and large disciplinary associations and journals bring together researchers from specialist or sub-disciplinary communities. Disciplines in this way still give structure to contemporary academia, be it in a more abstract sense. As a result, I expected to find quite extensive diffusion of publications when looking at individual author's research portfolios as well as the sharing of research subjects across disciplines.

<u>RQ3.1:</u> *Do authors who switch between disciplines throughout their careers change their research direction?* Disciplines do not have clearly defined boundaries; they are subject to change due to processes of internal differentiation and the emergence of inter-disciplines (Graff, 2015) or interdisciplinary specialties. I have presented a network representation based on discipline switches of authors to investigate along which routes (or on which intersections) interdisciplinary mobile groups could be present. The overall structure was slightly different from a network representation based on the cognitive similarities between disciplines, suggesting that synergies not only happen between disciplines working on similar

topics or subjects. Additionally, I also expected to find an overall increase of disciplinary boundary crossing occurrences over the course of a researcher's career. While the ratio of discipline switches is indeed steadily increasing, the ratio of disciplinary broadening remains stable.

RQ3.2: *And how does this relate to the distance they have travelled cognitively speaking?* At the same time, I found that the number of discipline switches an author makes throughout his/her career does not necessarily mean that he or she is conducting 'new' work in terms of changing topic or subject scope. In other words, it was found that the cognitive distance traveled by an author does not relate to the number of discipline switches he or she makes. The contrary was true for broadening occurrences. Authors who more frequently broadened their disciplinary horizon were found to be cognitively more mobile than those who only rarely visited new disciplines. As it was found that the vast majority of researchers switch between disciplines at least once (+65 %) or visit new disciplines (+75%) it becomes relevant to ask RQ3.2.

RQ3.3: *Which subjects are shared between disciplines and which disciplines are more open to sharing subjects with other disciplines?* In the second study presented in part 3 (chapter 8). we have therefore conducted a text clustering analysis to approximate subject specialties in VABB-SHW. On the level of disciplines (in terms of the disciplinary affiliation of researchers and the cognitive classification of the records authored by researchers), it was found that they all share subjects with one another, but some disciplines can be thought of as being more 'open' than others, i.e., they are less dominant within the subject specialties they share with other disciplines. Examples of such open disciplines are: Educational sciences, History, Sociology, and Media and communications. Closed fields include Law, Languages, Religion, and Economics and Business. When looking at the alignment of the dominant discipline classifications present within subject specialties, we found that these two types of classifications were generally in agreement, but some cross-overs were found as well.

As it was found that all subject specialties identified exhibited a relatively high disciplinary diversity, I wanted to understand if a typology could be discerned. This led to the final research question.

<u>RQ3.4:</u> *Can we find different types of specialties in terms of their disciplinary identities?* At first sight, the results looked messy, but when I contrasted the diversity in terms of cognitive and organizational disciplinary classifications present within the clusters, a linear relationship (with high heteroskedasticity) became apparent. The general trend seemed to be: a higher organizational diversity relates to higher cognitive diversity. The scatterplot could be divided into quartiles, and after inspection each of these quartiles was labeled. Four types of subject specialties were identified, ranging from disciplined and institutionalized to fractured cross-disciplinary specialties.

9.2 Implications for scientometric research

Both disciplines and interdisciplinarity are 'moving targets' (Ash, 2019). In fact, these were also some of the early warnings I received when I was visiting my first conference in bibliometrics (STI2018) and informed others about the topic of my research: "Are you sure? It's a grey zone…" and "Know what you are getting into, it's a mess". While I am now able to say that I totally agree with these statements, I do think that some things became clearer along the way. (1) Interdisciplinarity is not new; it emerged (and continuously emerges) together with disciplines and specialties, which makes the concept hard to grasp and define. This is why we need a historically informed and bottom-up approach to discipline and specialty formation. From another angle, (2) complementing a disciplinary ontology could be useful in discerning the multitude of interactions which happen between fields and other institutions in society than the university. As pointed out by Ash, Weingart and others, transdisciplinary connections are becoming ever more relevant. Classifying research in terms of disciplines is not entirely relevant in such a context.

The implications of these findings are important for scientometric research, as they show that differentiation along the lines of subjects of interest plays an increasingly important role in connecting the different disciplines and also other systems in society. It shows that, next to the traditional usage of broad discipline categories, subject specialties as cognitive communities should also be taken into consideration if we want to assess and understand interdisciplinarity (or transdisciplinarity) more systematically (Sjögårde, 2019). It should again be added that other lines of internal differentiation could apply, i.e. along methodological, theoretical, or epistemological.

As shown in the study on boundary crossing, disciplines should not be considered as isolated structures. They are strongly interdependent. In accordance with the literature on research specialties, the results show that specialization does not hinder interdisciplinarity. Specialization and interdisciplinarity are only in a seemingly paradoxical relationship (Weingart, 2000). Interdisciplinary but specialized knowledge communities exist at the boundaries of disciplines. In notable ways, these communities fulfill the functions (legitimation, certification and knowledge production) of disciplines to the extent that they along the way develop into discipline-like structures themselves.

With this in mind, I would like to argue that interdisciplinarity should again be studied 'from the ground up' – without solely relying on a flat disciplinary ontology. Research or subject specialties should be put to the foreground if we want to take interdisciplinarity more seriously. While this call for more systematic research into the structure and functioning of scientific specialties is almost as old as scientometrics and the sociology of science alike, it has not yet fully entered the contemporary debate on IDR, or at least not to the degree it should have. As pointed out in the literature overview of chapter 8, numerous case studies exist which point toward the interdisciplinary configurations present within subject specialties. As was shown in chapter 8, no single specialty is entirely disciplined, and all disciplines share subject ground with each other. Hence, it does not make much sense to consider disciplines as well delineated and stable categories.

I stress the importance of moving away from a purely disciplinary ontology when taking stock of the cognitive system of SSH, and yet I use these classifications myself. While I agree that such an approach conserves the methodology for which I am trying to propose solutions, such classifications are currently the best proxies of scholarly communities which are in place. Proposing a perspective in which subject specialties are considered a connecting component, crossing the so-called divides between disciplines and science and society might offer one of many possible ways forward.

9.3 Implications for research policy

Current research evaluation procedures focusing on interdisciplinarity are almost always centered around indicating the degree of interdisciplinarity of a particular unit of analysis. As we have discussed in chapters 1 and 3, these indicators rely quite heavily on diversity and/or coherence measurement for particular categories thought of as representative of (a part of) disciplinary knowledge. The indicators are of course informative. But as I have aimed to show, interdisciplinarity is omnipresent in the sciences at large, and also in the SSH in particular. IDR has emerged together with the disciplinary system and further intensifies together with the growth of that same system. A growing number of disciplines, sub-disciplines and specialist communities increases the number of possible interdisciplinary configurations.

The difficulty which we face in evaluating or indicating interdisciplinarity I believe for a large part has to do with a somewhat naïve assumption that we, as scientometricians, know how we should define and/or operationalize the knowledge communities or disciplines for which we want to assess the degree of interdisciplinarity. This assumption is partially reflected in, for example, a top-down assignment of categories to documents or units and the further fine-tuning of their granularity to better fit the methods or indicators which are being used. As I have also tried to argue, disciplines by default grow, further differentiate, decline, etc. and this is heavily dependent upon the broader socio-political context in which they exist. Additionally, in all fields, and on all levels within and between these fields, intense discussions are constantly going on about what the boundaries might be, and where potential lies ahead for broadening the horizon in terms of new research questions, methodologies, etc. A statistic developed from purely quantitative scientometric research indicating the degree of interdisciplinarity for a particular predefined field or set of categories, does not necessarily aid in moving forward in such debates, nor does it stimulate innovative interdisciplinary problem solving per se.

In other words, it is not the metrics and indicators that are problematic, but rather their application. More interesting and stimulating uses of the tools available within scientometrics to evaluate or indicate interdisciplinarity can be thought of when we think of research evaluation as a collaborative project or, as Rafols (2019) programmatically formulated this, as “indicators in the wild”. To create more meaningful insights about interdisciplinarity, scientometricians and science policy actors should (re-)engage with the social context for which these tools and indicators could be of relevance (i.e., fields, disciplines or specialisms). More clarity about what interdisciplinarity precisely entails or could entail at a point in time and within a specific social context could emerge from such collaborative and engaging acts. Such explorations and evaluations should start from the necessities which exist within these contexts.

What are the questions yet unsolved? A more valuable and constructive approach to research evaluation in the context of interdisciplinarity would be to apply the tools available in scientometrics to assess the diversity present within specialist communities. And what are possible (inter)disciplinary blind spots in the research carried out within these communities (Van Praag and Dhaenekindt, 2020)? What is needed in such a context are adequate mapping techniques, informed by the specialist communities which are being evaluated. What are the relevant literatures and other forms of communication? What are relevant stakeholders?

The role for evaluators in such a context is a facilitative one. It could for example consist of indicating and mapping research and other existing knowledge on a subject or problem of interest. Assessing the (inter-)disciplinary diversity already present within these communities by applying different diversity indicators and with multiple classification systems. Further stipulate the needs in terms of caveats in the knowledgebase, techniques or data, etc. Facilitate communication between different specialist communities which could be of relevance in solving an issue or addressing a particular theoretical/methodological need. Evaluate to what extent efforts are being made to communicate and co-create 'interdisciplinary' knowledge. To what extent are initiatives being taken to establish contacts between other specialist communities working on similar subject matter?

In short, it is naïve to assume that these indicators as standalone descriptors will help us in further enhancing or encouraging interdisciplinarity by simply measuring degrees of IDR. A more constructive and complexity sensitive approach would entail visiting the on-the-ground practices of researchers, consulting stakeholders and co-define their needs in terms of knowledge exchange. As Marres and de Rijcke (2020) and Rafols (2021) have recently stated, we should move from 'measuring' interdisciplinarity to indicating it. This process of 'indicating' takes us out of the 'statistical comfort zone' and requires scientometricians, policy makers and research evaluators to include other than purely quantitative and structural approaches when investigating interdisciplinary configurations or constructing policy guidelines.

9.4 Limitations

As a consequence of the broad scope of this thesis, I have only briefly summarized a general history of disciplines without paying particular attention to recent developments in the social sciences and humanities. My main goal here was to bring into attention insights from sociology and history of science to show that SSH disciplines are constantly reconfiguring. With regard to the aspect of internal differentiation, I recognize that my topical approach concerning subject specialties is limited. This is only one out of many possible axes along which knowledge communities establish themselves and further differentiate. The evidence presented in the literature does allow for some confidence in these generalist claims, although I recognize that devoted case-studies into interdisciplinarity in the SSH, or more systematic comparisons of the methods used with other types of variables indicative of specialty membership and/or cross-disciplinarity would be useful (see section 9.5).

The idea of a more encompassing knowledge society in which research is carried out and used in other sectors than higher education or academia is becoming ever more relevant. A perspective in which the academic system is considered as a part of a larger knowledge system or society at large (Etzkowitz, 2003), where exchange between other institutions than the university are taken into account, might in the future become even more necessary to adequately study cognitive or specialist communities.

Science in general is an increasingly international endeavor. While variations of the degree of internationalization exist across research domains in the SSH (Mosbah-Natanson and Gingras, 2014), it is hardly unthinkable a researcher today does not make use of sources outside of one’s own country. No single academic exclusively operates in a regional or national vacuum. The database used for our empirical studies, however, is regional by design. VABB-SHW only covers those publications which have been authored by (or when) researchers who are or were affiliated to a Flemish SSH research unit (Engels and Guns, 2018, p. 49-51). The comprehensive coverage of locally produced SSH scholarship it offers comes at the fairly expensive cost of ‘turning inward’. In our analyses we do not include the international milieu in which knowledge of particular subjects or disciplines is being created, yielding only a partial picture. It speaks for itself that this should be kept in mind when interpreting the results discussed above.

The use of content-based approaches has the advantage of being widely applicable to social sciences and humanities scholarship by avoiding asymmetries in citation and referencing practices across these fields. The negative implication is that no information on social or communicative relations inherent to knowledge production are taken into account. This is however, fundamental to develop a more adequate understanding of how cognitive communities operate. As suggested in the prospects for future research (section 9.5), including other bibliographic or sociometric variables would add value to the text-based methods applied in this thesis. In addition, I believe, the overtly structuralist perspective offered in this thesis hinders our understanding of how individual academics might be shaping and defining fields of research by acting interdisciplinary, but also how external actors and systems influence research practices (see Sile and Vanderstraeten, 2020). In general more interpretative research is needed in scientometrics to counteract the one-eyed structuralist dominance in the field (see Leydesdorff, Ràfols, & Milojević, 2020).

9.5 An agenda for future research

Automated document classification and clustering are extensively researched in scientometrics. Research specific to social sciences and humanities fields is largely lagging behind. Some preliminary steps were taken in this thesis to fill this gap, but much remains to be done. Interesting ways forward would include research on the evaluation of more recent text vectorization methods specifically for SSH documents. The techniques used in my research were either simple and established (i.e., BoW, TF-IDF, LSI, Word2Vec average, Doc2Vec) or more advanced and recent, but pre-trained embeddings models (i.e. Universal Sentence Encoders). A step forward would be to train neural network-based models like BERT on a large corpus of SSH literature. Research in other domains of science has shown that this increases the accuracy of word and document embedding models, and yields more qualitative results (biomedical BERT and legal BERT). The collection of an adequate multi-lingual corpus needed for training such a model, preferably consisting of full-text documents, would be a first step in that direction.

I have limited my explorations into automated document classification and clustering to textual data from titles and abstracts. While this is a highly generalizable framework, the inclusion of other variables and context sensitive evaluation metrics would likely be beneficial in terms of accuracy and textual coherence. A first possible expansion of the classification model developed in this thesis would be to include information about authors and/or the medium in which the publication appeared, etc. If two articles are authored by the same researcher, they are more likely to be related to the same subdiscipline or specialty (with exceptions, of course; co-authorship often brings researchers from far across their disciplinary or specialty turf together). Two articles appearing in the same journal are also more likely to be dealing with a similar subject of interest (again, limitations exist specifically for multidisciplinary journals). Another relevant step forward for classifying scientific documents would be, as I have stated in chapter 5, to take the cognitive similarity between categories into account when assessing the precision of classification outcomes (pp. 93-94).

It should also be noted that classification and clustering methods based on bibliographic information are not the only way to approach scientific communities on the level of subdisciplines or research specialties. Self-identification of researchers, (editorial board) membership of journals or conferences, or holding a particular academic degree has informed research into scientific specialties and disciplines as well. While these are not new approaches in themselves (see Mullins et al., 1977 for an example of such a comparison with co-citation clusters), to my knowledge no systematic comparative research exists in which advanced text clustering approaches are contrasted to these sociometric (e.g. membership or degree-based) methods. Such studies would not only allow us to assess their potential complementarity, they could also assist us in answering relevant questions posed by sociologists of science concerning the structure, development, and fragmentation of scientific specialties.

Disciplinary boundary crossing or mobility in the SSH, as studied in chapter 7, is more common than what one might assumed. We have identified numerous researchers who publish in or move between different discipline categories. A first interesting avenue for future work would be to include specialty membership in the analyses. As some subject specialties are indeed more disciplinarily diverse, it might well be that an author who crosses different disciplinary boundaries throughout his/her career is active within one or two interdisciplinary subject specialties. Another question relates to changes in one's specialist community of reference and his/her publication practices. I expect that those who move to more team-oriented specialties (on the border of medical sciences, for example) will most probably increase their productivity as well as their overall share of journal articles.

Chapter 8 has identified different types of disciplinary-interdisciplinary subject specialties. As touched upon in the discussion section of that chapter, from a scientometric perspective it would indeed be interesting to investigate the internal communication dynamics of these groups. A common assumption in the literature on interdisciplinarity is that this type of research is more impactful citation-wise, but most of these studies normalize citation counts by disciplinary categories. As indicated, investigations of inter-disciplinary diversity for citation patterns should be extended to intra-disciplinary differences as well. An assessment of the claim on increased impact could be done based on the operationalization of disciplinary-interdisciplinary specialties, be it in the context of other science domains. Citation practices largely differ across specialties, so a normalization not on the level of discipline categories but of specialties might reveal yet another dynamic at play.

10 Conclusions

The main conclusion which can be drawn from this thesis is that disciplinary specialization and interdisciplinarity emerge together and co-exist instead of being opposed or seemingly paradoxical. We have shown that in the SSH a variety of interdisciplinary fields exist, and that this is the norm rather than the exception. While the role played by different specialties in connecting the disciplines has been illustrated, it should also be kept in mind that the operationalization of these specialties is not straightforward. The classification studies presented, for example, detail the difficulty of reaching an agreement on specialty level categories for documents, both for experts and classification algorithms alike. Combinations of different mapping or categorization techniques can be more appropriate if we wish to further disentangle the complexity of the sciences.

The findings presented in this thesis can thus be thought of as a pointer for bibliometric research into specialization and interdisciplinarity. As I have tried to argue, increasing specialization leads to increased interdisciplinarity and this demands for a re-appraisal of specialty level bibliometric research. Such programmatic statements are not new. 20 years ago, Jochen Gläser for example issued a similar call, terming scientific specialties 'as the currently missing link between scientometrics and the sociology of science' (Gläser, 2001, p. 191). In retrospect, I have set out some preliminary endeavors into the direction of recombining the two, by for example studying variations in the cognitive structure and connecting this to institutional aspects. Suggestions were made about the degree of institutionalization and developmental stages of subject specialties, and I believe a great deal of potential lies ahead in further exploring this connection. Following the trajectory of emerging specialties in the social sciences and humanities from their conception and studying how they reconfigure on both the cognitive and institutional level would further our understanding of what facilitates and what hinders interdisciplinarity.

For research evaluation and policy concerned with interdisciplinary research, the key take-home lesson is that there exists a great diversity in disciplinary and interdisciplinary identities, both on the level of research communities or specialties and on the level of career trajectories of individual researchers. Implementing catch-all indicators that inevitably reduce this complexity do away with this diverse richness and run the risk of standardizing what interdisciplinarity *must* mean. The variety in possible interdisciplinarities, as the number of case-studies on interdisciplinary fields and our research indicate, is far too great to achieve consensus on one statistic or category being representative of all interdisciplinary practice. As such, we have proposed to foster co-creation models of evaluation in which we first situate and co-define interdisciplinary practice together with stakeholders involved in order to further encourage it. The role of research administrators and policymakers in this context would be one of building bridges.

Author contributions for co-authored pieces

Chapter 4. Fine-grained classification of social science journal articles using textual data: First steps

Joshua Eykens: Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Data curation, Writing—original draft, Writing—review & editing, Visualization, Supervision, Project administration.

Raf Guns: Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Writing—original draft, Writing—review & editing, Supervision.

Tim C.E. Engels: Conceptualization, Writing—review & editing, Supervision.

Chapter 5. Scaling up: Applying the system to multiple domains and multi-label documents

Joshua Eykens: Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Data curation, Writing—original draft, Writing—review & editing, Visualization, Supervision, Project administration.

Raf Guns: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing—original draft, Writing—review & editing, Visualization, Supervision.

Tim C.E. Engels: Conceptualization, Writing—original draft, Writing—review & editing, Supervision.

Chapter 6. Assessing different document vectorization techniques for unsupervised clustering

Joshua Eykens: Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Data curation, Writing—original draft, Writing—review & editing, Visualization, Supervision, Project administration.

Raf Guns: Conceptualization, Methodology, Software, Formal analysis, Writing—original draft, Writing—review & editing, Visualization, Supervision.

Tim C.E. Engels: Conceptualization, Writing—original draft, Writing—review & editing, Supervision.

Chapter 7. Crossing disciplinary boundaries?: Cognitive and disciplinary mobility in the social sciences and humanities

Joshua Eykens: Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Data curation, Writing—original draft, Writing—review & editing, Visualization, Supervision, Project administration.

Raf Guns: Methodology, Software, Data curation, Writing—review & editing.

Tim Engels: Conceptualization, Methodology, Writing—review & editing, Supervision.

Chapter 8. Subject specialties as interdisciplinary trading grounds: The case of the social sciences and humanities

Joshua Eykens: Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Data curation, Writing—original draft, Writing—review & editing, Visualization, Supervision, Project administration.

Raf Guns: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing—review & editing, Visualization, Supervision.

Raf Vanderstraeten: Conceptualization, Methodology, Investigation, Writing—original draft, Writing—review & editing.

Writings and talks prepared during PhD

Articles in peer reviewed journals

Eykens, J., Guns, R., & Engels, T.C.E. Crossing disciplinary boundaries? Interdisciplinary mobility in the social sciences and humanities. Manuscript in preparation.

Eykens, J. Guns, R., & Vanderstraeten, R. (2022). Subject specialties as interdisciplinary trading grounds: The case of the social sciences and humanities. *Scientometrics, 128*. doi: https://doi.org/10.1007/s11192-021-04254-w

Teixeira da Silva, J. A., Dunleavy, D. J., Moradzadeh, M., & Eykens, J. (2021). A credit-like rating system to determine the legitimacy of scientific journals and publishers. *Scientometrics*, 126, 8589-8616. doi: https://doi.org/10.1007/s11192-021-04118-3

Eykens, J. Guns, R., & Engels, T. C. E. (2021). Fine-grained classification of social science journal articles using textual data: a comparison of supervised machine learning approaches. *Quantitative Science Studies, 2020*, 1-26. doi: https://doi.org/10.1162/QSS_A_00106

Eykens, J., Guns, R., & Engels, T. C. E. (2020). Questionable journals on a hunt: Findings from research into a national bibliographic database (in Dutch*). TH&MA: Tijdschrift voor Hoger Onderwijs en Management*, 3, 90-95. https://repository.uantwerpen.be/docstore/d:irua:1284

Eykens, J., Guns, R., Puuska, H., Pölönen, J., & Engels, T. C. E. (2020). A shared European research space for the social sciences and humanities? English language publishing and the use of European journals. *Darbai ir Dienos, 73*, 37-51. doi: https://doi.org/10.7220/2335-8769.73.2

Eykens, J., Guns, R., Rahman, A. J., & Engels, T. C. E. (2019). Identifying publications in questionable journals in the context of performance-based research funding. *PloS one*, *14*(11), e0224541. doi: https://doi.org/10.1371/journal.pone.0224541

Guns, R., Eykens, J., & Engels, T. C. (2019). To what extent do successive cohorts adopt different publication patterns? Peer review, language use, and publication types in the social sciences and humanities. *Frontiers in research metrics and analytics, 3*, 1-16. doi: https://doi.org/10.3389/FRMA.2018.00038

Guns, R., Sile, L., Eykens, J., Verleysen, F., & Engels, T. C. E. (2018). A comparison of cognitive and organizational classification. *Scientometrics, 116*(2), 1093-1111. doi: https://doi.org/10.1007/S11192-018-2775-X

Code and data

Eykens, J., & Guns, R. (2021). interdisciplinarity-subject-specialties-SSH. https://github.com/JoshuaE1/interdisciplinarity-subject-specialties-SSH

Eykens, J., & Guns, R. (2021). Interdisciplinary subject specialties SSH (Version 1). Zenodo. doi: https://doi.org/10.5281/zenodo.6337574

Eykens, J., & Guns, R. (2020). Supervised classification of SSH publications. *Zenodo*. doi: https://doi.org/10.5281/zenodo.382230

Eykens, J., & Guns, R. (2019). Identifying publications in questionable journals in the context of PRFS: code and data (v 1.0). *Zenodo*. doi: https://doi.org/10.5281/zenodo.2842559

Articles in magazines

Van Beek, F. & Eykens, J. (2020). Machine learning voor sociologen (Dutch). *Sociologie Magazine, 28*(2), 20-22.

Eykens, J. (2019). Wetenschappers in actie: een gesprek met Frederic Vandermoere and Ilse Loots (Dutch). *Sociologie Magazine, 27*(2), 20-21.

Eykens, J. (2018). Mobility within academia: A conversation with Cassidy R. Sugimoto and Vincent Larivière (Dutch). *Sociologie Magazine, 26*(4), 22-24.

Vanderstraeten, R., Louckx, K., & Eykens, J. (2018). Sociology in Belgium: A conversation with Raf Vanderstraeten and Kaat Louckx (in Dutch). *Sociologie Magazine, 26*(3), 22-23.

Eykens, J. (2018). Editorial: Energy (in Dutch). *Sociologie Magazine, 26*(3), 2.

Eykens, J. (2018). Big data voor sociologen (in Dutch). *Sociologie Magazine, 26*(2), 30-31.

Book chapter

Eykens, J. (2022). Assessing interdisciplinary research in the social sciences: Are we on the right track? In Engels, T.C.E. and Kulczycki, E. (Eds.), *Handbook on Research Assessment in the Social Sciences*. London: Edward Elgar Publishing.

Conference proceedings and posters

Eykens, J., Guns, R., & Vanderstraeten, R. (2021). Subject specialties as interdisciplinary trading grounds: The case of the social sciences and humanities. Poster presented at the Nordic Workshop on Bibliometrics and Research Policy, Odense, Denmark.

Eykens, J., Guns, R., & Engels, T. C. E. (2021). Clustering social sciences and humanities publications: Can word and document embeddings improve cluster quality? Manuscript submitted as research in progress paper for ISSI 2021.

Sīle, L., Aspeslagh, P., Eykens, J., & Guns, R. (2020). National bibliographic data for studies of social sciences and humanities: Towards interoperability. ITM Web of Conferences, 33, 02002. https://doi.org/10.1051/itmconf/20203302002

Eykens, J., Guns, R., & Engels, T. C. E. (2019). Article level classification of publications in sociology : an experimental assessment of supervised machine learning approaches. In Catalano, G. et al. (eds.), Proceedings of the 17th international conference on scientometrics & informetrics (ISSI 2019), 2-5 September, 2019, Sapienza University of Rome, Italy (Vol. 1, pp. 738–743). Rome: Edizioni Efesto.

Eykens, J., Guns, R., Rahman, A. I. M. J., Sīle, L., & Engels, T. C. E. (2018). Predatory Open Access journals: A review of past screenings within the Flemish performance based research funding system (2014 – 2018). In STI 2018 Conference Proceedings (pp. 799–806). Leiden: Centre for Science and Technology Studies (CWTS).

Reports

Guns, R., Arhiliuc, C., Eykens, J., & Vandewalle, E. (2022). Screening VABB-SHW version 12 for publications in predatory journals: report to the authoritative panel.

Eykens, J., Guns, R., & Zhou, H. (2021). Assessing interdisciplinary research involving the social sciences and humanities. *Vlaams indicatorenboek: wetenschap, technologie en innovatie, 2021*

Eykens, J., & Guns, R. (2020). Screening VABB-SHW version X with Cabells Journal Blacklist and Directory of Open Access Journals: report to the authoritative panel. https://hdl.handle.net/10067/1674140151162165141

Guns, R., Eykens, J., & Engels, T. (2019). Publications in questionable journals. In Debackere, K. & Veugelers, R. (eds.), *Vlaams indicatorenboek: wetenschap, technologie en innovatie, 2019* (pp. 286–290). Leuven: Departement EWI & ECOOM.

Eykens, J. Guns, R., & Engels, T. C. E. (2018). Comparison of VABB-SHW version IX with Cabells Journal Blacklist and Directory of Open Access Journals : report to the Authoritative Panel. https://hdl.handle.net/10067/1571090151162165141

Eykens, J. Guns, R., & Engels, T. C. E. (2018). Comparing VABB-SHW (version VIII) with Cabells Journal Blacklist and Directory of Open Access Journals : report to the Authoritative Panel. https://hdl.handle.net/10067/1481030151162165141

Talks and presentations during PhD

Eykens, J. (2021). Subject specialties as interdisciplinary trading grounds: The case of the social sciences and humanities. Poster presentation at the Nordic Workshop on Bibliometrics and Research Policy, Odense, Denmark.

Eykens, J. (2021). How to approach interdisciplinarity in the social sciences and humanities? Presentation at ISSI 2021, virtual conference, Leuven, Belgium.

Eykens, J. (2021). Clustering social sciences and humanities publications: Can word and document embeddings improve cluster quality? Presentation at ISSI 2021, virtual conference, Leuven, Belgium.

Eykens, J. (February 21st, 2020). Screening the Flemish Academic Bibliographic Database for the SS&H (VABB-SHW) for publications in questionable journals. Invited talk at 'Changing publication practices – From predatory to Plan S' organized by the Young Academy and the Royal Flemish Academy of Belgium for Science and the Arts (KVAB), Palace of the Academies, Brussels, Belgium.

Eykens, J. (2019). Article level classification of publications in sociology: An experimental assessment of supervised machine learning approaches. Presentation at ISSI 2019, Rome, Italy.

Eykens, J. (November 26th, 2018). Disciplinarity and interdisciplinarity in the social sciences and humanities: Some considerations and a proposal for further research. Presentation at the internal ECOOM-day, Leuven, Belgium.

Eykens, J. (September 12th, 2018). Predatory Open Access journals: A review of past screenings within the Flemish Performance Based

Research funding system (2014 – 2018). Presentation at STI – Science and Technology Indicators 2018, Leiden, the Netherlands.

Eykens, J. (July 5th, 2018). A situated understanding of media ecologies and media practices in social movement studies: Network ethnography as a methodological lens. Presentation at NetGloW – Networks in a Global World 2018, Saint Petersburg, Russia.

Eykens, J. (May 25th, 2018). The PRFS and VABB-SHW: Challenges with the selection of books and book publications. Presentation at ENRESSH meeting, Madrid, Spain.

Bibliography

Abbott, A. (2001). *Chaos of disciplines*. Chicago: The University of Chicago Press.

Abma, R. (2011). *Over de grenzen van disciplines: plaatsbepaling van de sociale wetenschappen*. Nijmegen: Uitgeverij Vantilt.

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2012). Identifying interdisciplinarity through the disciplinary classification of coauthors of scientific publications. *Journal of the American Society for Information Science and Technology, 63*(11), 2206-2222. doi:https://doi.org/10.1002/asi.22647

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. doi:https://doi.org/10.1016/j.joi.2018.09.001

Allmendinger, J. (2015). *Quests for interdisciplinarity: A challenge for the ERA and HORIZON 2020*. Retrieved from: https://ec.europa.eu/info/sites/default/files/research_and_innovation/groups/rise/allmendinger-interdisciplinarity.pdf

Amir, S. (1985). On the degree of interdisciplinarity of research programs: A quantitative assessment. *Scientometrics, 8*(1-2), 117-136. doi:https://doi.org/10.1007/bf02025225

Angelov, D. (2020). Top2Vec: Distributed Representations of Topics. *Arxiv preprint arXiv:2008:09470*.

Aris, S. (2021). Fragmenting and connecting? The diverging geometries and extents of IR's interdisciplinary knowledge-relations. *European Journal of International Relations, 27*(1), 175-203. doi:https://doi.org/10.1177/1354066120922605

Ash, M. G. (2019). Interdisciplinarity in Historical Perspective. *Perspectives on Science, 27*(4), 619-642. doi: https://doi.org/10.1162/posc_a_00318

Bammer, G. (2016). What Constitutes Appropriate Peer Review for Interdisciplinary Research. *Palgrave Communications, 2*. doi:http://dx.doi.org/10.1057/palcomms.2016.17

Becher, P., & Trowler, R. (2001). *Academic Tribes and Territories: Intellectual enquiry and the culture of disciplines*: The Society for Research into Higher Education and Open University Press.

Bergmann, M., Brohmann, B., Hoffmann, E., Loibl, C., Rehaag, R., Schramm, E., & Vos, J. (2005). Quality Criteria of Transdisciplinary Research: A guideline for the fromative evaluation of research projects. *ISOE-Studientexte*, *13*.

Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., & Cox, D. D. (2015). Hyperopt: a Python library for model selection and hyperparameter optimization. *Computational Science & Discovery, 8*(1). doi:https://doi.org/10.1088/1749-4699/8/1/014008

Blaemers, J. (2006). A History of the Development and Maintenance of the Sociological Abstracts Thesaurus of Sociological Indexing Terms. *Behavioral & Social Sciences Librarian, 25*(1), 111-122. doi:https://doi.org/10.1300/J103v25n01_05

Boix-Mansilla, V. (2006). Assessing expert interdisciplinary work at the frontier: an empirical exploration. *Research Evaluation, 15*(1), 17-29. doi:https://doi.org/10.3152/147154406781776075

Bonaccorsi, A. (2022). 3 An epistemic approach to research assessment in the social sciences. In T. Engels & E. Kulczycki (Eds.), *Handbook on Research Assessment in the Social Sciences* (pp. 14-47). UK: Edward Elgar Publishing.

Booth, B. (1996). *Thesaurus of Sociological Indexing Terms*. The University of Michigan: Sociological Abstracts, Inc.

Bornmann, L., Haunschild, R., & Mutz, R. (2021). Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanities and Social Sciences Communications, 8*(224). doi:https://doi.org/10.1057/s41599-021-00903-w

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. doi: https://doi.org/10.1016/j.joi.2014.04.001

Boyack, K. W., Newman, D., Duhon, R. J., Klavans, R., Patek, M., Biberstine, J. R., . . . Börner, K. (2011). Clustering More than Two Million Biomedical Publications: Comparing the Accuracies of Nine Text-Based Similarity Approaches. *PLOS ONE, 6*(3), e18029. doi:https://doi.org/10.1371/journal.pone.0018029

Bromham, L., Dinnage, R., & Hua, X. (2016). Interdisciplinary research has consistently lower funding success. *Nature, 534*(7609), 684-687. doi:https://doi.org/10.1038/nature18315

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 chemistry. *Scientometrics, 22*(1), 155-205. doi: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. doi:10.1177/053901883022002003

Campbell, D. T. (1969). Ethnocentrism of disciplines and the fish-scale model of omniscience. In C. Sherif & W. Sherif (Eds.), *Interdisciplinary Relationships in the Social Sciences*: Aldine.

Cer, D., Yang, Y., Kong, S.-y., Hua, N., Limtiaco, N., John, R. S., . . . Kurzweil, R. (2018). Universal Sentence Encoder. *arXiv preprint ArXiv:1803.11175v2*.

Chakraborty, T., Tammana, V., Ganguly, N., & Mukherjee, A. (2015). Understanding and modeling diverse scientific careers of researchers. *Journal of Informetrics, 9*(1), 69-78. doi:https://doi.org/10.1016/j.joi.2014.11.008

Chubin, D. E. (1976). State of the Field The Conceptualization of Scientific Specialties. *The Sociological Quarterly, 17*(4), 448-476. doi:https://doi.org/10.1111/j.1533-8525.1976.tb01715.x

Clark, S. G., & Wallace, R. L. (2015). Integration and interdisciplinarity: concepts, frameworks, and education. *Policy Sciences, 48*(2), 233-255. doi:https://doi.org/10.1007/s11077-015-9210-4

Clauset, A., Newman, M. E. J., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E, 70*(6), 066111. doi:https://doi.org/10.1103/PhysRevE.70.066111

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., . . . Helbing, D. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics, 214*(1), 325-346. doi:https://doi.org/10.1140/epjst/e2012-01697-8

De Domenico, M., Omodei, E., & Arenas, A. (2016). Quantifying the diaspora of knowledge in the last century. *Applied Network Science, 1*(1), 15. doi:https://doi.org/10.1007/s41109-016-0017-9

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6). doi:https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Dolgin, E. (2021). The tangled history of mRNA vaccines. *Nature, 597*(7876), 318-324. doi:https://doi.org/10.1038/d41586-021-02483-w

Dunham, J., Melot, J., & Murdick, D. (2020). Identifying the Development and Application of Artificial Intelligence in Scientific Text. *arXiv preprint arXiv:2002.07143.*

Durkheim, E. (2014). *The division of labor in society*. New York, N.Y.: Free Press.

Egghe, L., & Rousseau, R. (2003). A measure for the cohesion of weighted networks. *Journal of the American Society for Information Science and Technology, 54*(3), 193-202. doi: https://doi.org/10.1002/asi.10155

Engels, T. C. E., & Guns, R. (2018a). The Flemish Performance-based Research Funding System: A Unique Variant of the Norwegian Model. *Journal of Data and Information Science, 3*(4), 45-60. doi:https://doi.org/10.2478/jdis-2018-0020

Engels, T. C. E., Ossenblok, T. L. B., & Spruyt, E. H. J. (2012). Changing publication patterns in the Social Sciences and

Humanities, 2000–2009. *Scientometrics, 93*(2), 373-390. doi:https://doi.org/10.1007/s11192-012-0680-2

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In E. Simoudis, J. Han, & U. Fayyad (Eds.), *Proceedings of the second international conference on knowledge discovery and data mining, August 2–4, Portland, Oregon, USA* (pp. 226-231): Association for the Advancement of Artificial Intelligence.

Science Europe. (2019). *Symposium Report: Interdisciplinarity*. Retrieved from: http://scieur.org/interdisc-symp

Etzkowitz, H. (2003). Innovation in Innovation: The Triple Helix of University-Government Relations. *Social Science Information, 42*(3), 293-337. doi:https://doi.org/10.1177/05390184030423002

Evans, E. D. (2016). Measuring Interdisciplinarity Using Text. *Socius: Sociological Research for a Dynamic World, 2*, 1-18. doi: https://doi.org/10.1177/2378023116654147--

Eykens, J., & Guns, R. (2020). Supervised classification of SSH publications. doi:https://doi.org/10.5281/zenodo.3822309

Eykens, J., Guns, R., & Engels, T. C. E. (2019). Article level classification of publications in sociology: An experimental assessment of supervised machine learning approaches. In G. Catalano, C. Daraio, M. Gregori, H. F. Moed, & G. Ruocco (Eds.), *17th international conference on scientometrics & informetrics (ISSI2019)* (Vol. 1, pp. 738-743). Sapienza University of Rome, Italy: Edizioni Efesto.

Eykens, J., Guns, R., & Engels, T. C. E. (2021). Clustering social sciences and humanities publications : can word and document embeddings improve cluster quality? . In W. Glänzel, S. Heeffer, P.-S. Chi, & R. Rousseau (Eds.), *Proceedings of the 18th conference of the International Society for Scientometrics and Informetrics* (pp. 369-374). Leuven, Belgium International Society for Scientometrics and Informetrics.

Ferreira, M. R., & Costas, R. (2021). Doing as the Romans do? A comparative analysis of the thematic profile of newcomers and already existing faculty at universities. In W. Glänzel, S. Heeffer, P.-S. Chi, & R. Rousseau (Eds.), *Proceedings of the 18th International Conference on Scientometrics and Informetrics - ISSI2021* (pp. 401-406). KU Leuven: International Society for Scientometrics and Informetrics (ISSI).

Fontana, M., Iori, M., Montobbio, F., & Sinatra, R. (2020). New and atypical combinations: An assessment of novelty and interdisciplinarity. *Research Policy, 49*(7). doi: https://doi.org/10.1016/j.respol.2020.104063

Funk, M. E. R., C. A. (1983). Indexing consistency in MEDLINE. *Bulletin of the Medical Library Association, 71*(2), 176-183. Retrieved from https://pubmed.ncbi.nlm.nih.gov/6344946/

Gieryn, T. F. (1983). Boundary-Work and the Demarcation of Science from Non-Science: Strains and interests in Professional Ideologies of Scientists. *American Sociological Review, 48*(6), 781-795.

Glänzel, W., & Debackere, K. (2021). Various aspects of interdisciplinarity in research and how to quantify and measure those. *Scientometrics*. doi:https://doi.org/10.1007/s11192-021-04133-4

Glänzel, W., Schubert, A., & Czerwon, H.-J. (1999). An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. *Scientometrics, 44*(3), 427-439. doi:https://doi.org/10.1007/BF02458488

Gläser, J. (2001). Scientific Specialties as the (Currently Missing) Link Between Scientometrics and the Sociology of Science. in Proceedings of the Eighth International Conference on Scientometrics and Informetrics, Sydney, Australia, 1, pages 191–210.

Goldstone, R. L., & Leydesdorff, L. (2006). The Import and Export of Cognitive Science. *Cognitive Science, 30*(6), 983-993. doi: https://doi.org/10.1207/s15516709cog0000_96

Gowanlock, M., & Gazan, R. (2013). Assessing researcher interdisciplinarity: a case study of the University of Hawaii NASA Astrobiology Institute. *Scientometrics, 94*, 133-161. doi:https://doi.org/10.1007/s11192-012-0765-y

Graf, J. (2019). *Bringing Concepts Together: Interdisciplinarity, Transdisciplinarity, and SSH Integration*. Paper presented at the fteval Journal for Research and Technology Policy Evaluation Issue 48/July 2019 - Proceedings of the Conference "Impact of Social Sciences and Humanities for a European Research Agenda Valuation of SSH in mission-oriented research".

Graff, H. J. (2015). *Undisciplining Knowledge: Interdisciplinarity in the Twentieth Century*. Baltimore, Maryland: Johns Hopkins University Press.

Graff, H. J. (2016). The "Problem" of Interdisciplinarity in Theory, Practice, and History. *Social Science History, 40*(4), 775-803. doi:https://doi.org/10.1017/ssh.2016.31

Guetzkow, J., Lamont, M., & Mallard, G. (2004). What is Originality in the Humanities and the Social Sciences. *American Sociological Review, 69*, 190-212.

Guns, R., Sīle, L., Eykens, J., Verleysen, F. T., & Engels, T. C. E. (2018). A comparison of cognitive and organizational classification of publications in the social sciences and humanities. *Scientometrics, 116*(2), 1093-1111. doi: https://doi.org/10.1007/s11192-018-2775-x

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. In G. Varoquaux, T. Vaught, & J. Millman (Eds.), Proceeings of the 7Th Python in Science Conference (SciPy2008) (pp. 11-15). Pasadena, CA, USA.

Hammarfelt, B. (2018). What is a discipline? The conceptualization of research areas and their operationalization in bibliometric research. In R. Costas, T. Franssen, & A. Yegros-Yegros (Eds.), *Science, Technology and Innovation Indicators in Transition - STI2018* (pp. 197-203). Leiden, The Netherlands: Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands

Hammarfelt, B. (2019). Discipline. In B. Hjørland & C. Gnoli (Eds.), *ISKO Encyclopedia of Knowledge Organization*. Toronto, Canada: International Society for Knowledge Organization.

Hargens, L. L. (1986). Migration patterns of U. S. Ph. D. s among disciplines and specialties. *Scientometrics, 9*(3), 145-164. doi: https://doi.org/10.1007/BF02017238

Heilbron, J. (1990). *Het onstaan van de sociologie*. Amsterdam: Prometheus.

Hicks, D. (2005). The Four Literatures of Social Science. In *Handbook of Quantitative Science and Technology Research* (pp. 473-496).

Honnibal, M., & Montani, I. (2018). spaCy 2.0.11. doi:https://doi.org/10.5281/zenodo.4291179

Huutoniemi, K. (2010). Interdisciplinarity in research evaluation. In R. Frodeman, J. T. Klein, & C. Mitcham (Eds.), *The Oxford Handbook of Interdisciplinarity* (pp. 309–320): Oxford University Press.

Huutoniemi, K., Klein, J. T., Bruun, H., & Hukkinen, J. (2010). Analyzing Interdisciplinarity: Typology and Indicators. *Research Policy, 39*(1), 79-88. doi:https://doi.org/10.1016/j.respol.2009.09.011

Huutoniemi, K., & Rafols, I. (2017). Interdisciplinarity in Research Evaluation. In R. Frodeman, J. T. A. E. Klein, & R. C. S. A. E. Pacheco (Eds.), *The Oxford Handbook of Interdisciplinarity* (Vol. 2nd, pp. 498-512). Oxford, United Kingdom: Oxford University Press.

Jacobs, J. (2013). *In defense of disciplines : interdisciplinarity and specialization in the research university*. Chicago: The University of Chicago Press.

Jacobs, J., & Frickel, S. (2009). Interdisciplinarity: A Critical Assessment. *Annual Review of Sociology, 35*(1), 43-65. doi: https://doi.org/10.1146/annurev-soc-070308-115954

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

Jost, L. (2009). Mismeasuring biological diversity: Response to Hoffmann and Hoffmann (2008). *Ecological Economics, 68*(4), 925-928. doi: https://doi.org/dx.doi.org/10.1016/j.ecolecon.2008.10.015

Kandimalla, B., Rohatgi, S., Wu, J., & Lee Giles, C. (2020). Large Scale Subject Category Classification of Scholarly Papers with Deep Attentive Neural Networks. *arXiv preprint arXiv:2007.13826.*

Karunan, K., Lathabai, H. H., & Prabhakaran, T. (2017). Discovering interdisciplinary interactions between two research fields using citation networks. *Scientometrics, 113*(1), 335-367. doi:https://doi.org/10.1007/s11192-017-2481-0

Katzner, D., W. (2003). Why mathematics in economics? *Journal of Post Keynesian Economics, 25*(4), 561-574. doi: https://doi.org/10.1080/01603477.2003.11051373

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Neural Information Processing Systems 2017* (pp. 1-9). Long Beach, California: Curran Associates, Inc.

Keynejad, R. C., Yapa, H. M., & Ganguli, P. (2021). Achieving the sustainable development goals: investing in early career interdisciplinarity. *Humanities and Social Sciences Communications, 8*(1). doi:https://doi.org/10.1057/s41599-021-00834-6

Klein, J. T. (2004). Guiding questions for integration. In *Proceedings of the Integration Symposium 2004*.

Klein, J. T. (2008). Evaluation of Interdisciplinary and Transdisciplinary Research. *American Journal of Preventive Medicine, 38*, S116-S123. doi:https://doi.org/10.1016/j.amepre.2008.05.010

Klima, N., Meysman, S., Carlier, J., Dewaele, A., & de Smet, E. (2019). *Ghent University's interdisciplinary SSH-consortia – A strategy to enhance the societal impact of research*. Paper presented at the fteval Journal for Research and Technology Policy Evaluation Issue 48/July 2019 - Proceedings of the

Conference "Impact of Social Sciences and Humanities for a European Research Agenda Valuation of SSH in mission-oriented research".

Kulczycki, E., Engels, T. C. E., Pölönen, J., Bruun, K., Dušková, M., Guns, R., . . . Zuccala, A. (2018). Publication patterns in the social sciences and humanities: Evidence from eight European countries. *Scientometrics, 116*, 463-486. doi: https://doi.org/10.1007/s11192-018-2711-0

Lamont, M. (2009). *How professors think: Inside the curious world of academic judgement*. Cambridge, MA: Harvard University Press.

Lamont, M., & Guetzkow, J. (2016). How quality is recognized by peer review panels: The case of the humanities. In M. Ochsner & S. E. Hug (Eds.), *Research Assement in the Humanities: Towards Criteria and Procedures* (pp. 31-41). Switzerland: Springer, Cham.

Langlois, A., Nie, J. Y., Thomas, J., Hong, Q. N., & Pluye, P. (2018). Discriminating between empirical studies and nonempirical works using automated text classification. *Research Synthesis Methods, 9*(4), 587-601. doi:https://doi.org/10.1002/jrsm.1317

Larivière, V., Archambault, E., Gingras, Y., & Vignola-Gagné, E. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology, 57*(8), 997-1004. doi:https://doi.org/10.1002/asi.20349

Larivière, V., & Gingras, Y. (2010). On the relationship between interdisciplinarity and scientific impact. *Journal of the American Society for Information Science and Technology, 61*(1), 126-131. doi:https://doi.org/10.1002/asi.21226

Larivière, V., & Gingras, Y. (2014). Measuring Interdisciplinarity. In B. Cronin & C. R. Sugimoto (Eds.), *Beyond Bibliometrics: Harnessing Multidimensional Indicators for Scholarly Impact* (pp. 187-200). Cambridge MA: MIT Press.

Laudel, G. (2006). Conclave in the Tower of Babel: How peers review interdisciplinary research proposals. *Research Evaluation, 15*(1), 57-68.

Le Pair, C. (1980). Switching between academic disciplines in universities in the Netherlands. *Scientometrics, 2*(3), 177-191. doi:https://doi.org/10.1007/BF02016696

Le, Q., & Mikolov, T. (2014). Distributed Representations of Sentences and Documents. In *Proceedings of the 31st International Conference on Machine Learning* (Vol. 32). Beijing, China: JMLR: W&CP.

Ledford, H. (2015). How to solve the world's biggest problems. *Nature, 525*(7569), 308-311. doi:https://doi.org/10.1038/525308a

Leininger, K. (2000). Interindexer consistency in PsycINFO. *Journal of Librarianship and Information Science, 32*(1), 4-8. doi: https://doi.org/10.1177/096100060003200102

Lelu, A., & Cadot, M. (2019). Evaluation of text clustering methods and their dataspace embeddings: an exploration. In *IFCS 2019 - 16th International on the Federation of Classification Societies* (pp. 1-24). Thessaloniki, Greece.

Levitt, J. M., Thelwall, M., & Oppenheim, C. (2011). Variations between subjects in the extent to which the social sciences have become more interdisciplinary. *Journal of the American Society for Information Science and Technology, 62*(6), 1118-1129. doi:https://doi.org/10.1002/asi.21539

Lewis, D. D. (1998). Naïve (Bayes) at forty: The independence assumption in information retrieval. In C. Nédellec & C. Rouveirol (Eds.), *10th European Conference on Machine Learning - ECML-98.* (Vol. 1398, pp. 4-15). Chemnitz, Germany: Springer, Berlin, Heidelberg.

Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology, 58*(9), 1303-1319. doi:https://doi.org/10.1002/asi.20614

Leydesdorff, L., & Goldstone, R. L. (2013). Interdisciplinarity at the journal and specialty level: The changing knowledge bases of the journal of cognitive science. *Journal of the Association for Information Science and Technology, 65*(1), 164-177. doi: https://doi.org/10.1002/asi.22953

Leydesdorff, L., & Ivanova, I. (2020). The measurement of "interdisciplinarity" and "synergy" in scientific and extra-scientific collaborations. *Journal of the Association for Information Science and Technology, 72*(4), 387-402. doi: https://doi.org/10.1002/asi.24416

Leydesdorff, L., & Probst, C. (2009). The delineation of an interdisciplinary specialty in terms of a journal set: The case of communication studies. *Journal of the American Society for Information Science and Technology, 60*(8), 1710-1718. doi:https://doi.org/10.1002/asi.21052

Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2018). Betweenness and diversity in journal citation networks as measures of interdisciplinarity. A tribute to Eugene Garfield. *Scientometrics, 114*(2), 567-592. doi:https://doi.org/10.1007/s11192-017-2528-2

Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2019). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics, 13*(1), 255-269. doi:https://doi.org/10.1016/j.joi.2018.12.006

Light, R., & Adams, J. (2017). A Dynamic, Multidimensional Approach to Knowledge Production. In S. Frickel, M. Albert, & B. Prainsack (Eds.), *Investigating Interdisciplinary Collaboration: Theory and Practice across Disciplines* (pp. 127-147). New Brunswick, New Jersey: Rutgers University Press.

Liu, M., Hu, X., & Li, J. (2018). Knowledge flow in China's humanities and social sciences. *Quality & Quantity, 52*(2), 607-626. doi:https://doi.org/10.1007/s11135-017-0539-y

Liu, M., Shi, D., & Li, J. (2017). Double-edged sword of interdisciplinary knowledge flow from hard sciences to humanities and social sciences: Evidence from China. *PLOS ONE, 12*(9), e0184977. doi:https://doi.org/10.1371/journal.pone.0184977

Liu, Y., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion. *J. Assoc. Inf. Sci. Technol., 61*, 340-351.

Lockett, A., & McWilliams, A. (2005). The Balance of Trade Between Disciplines: Do We Effectively Manage Knowledge? *Journal of Management Inquiry, 14*(2), 139-150. doi:https://doi.org/10.1177/1056492605276645

Loper, E., & Bird, S. (2002). NLTK: the Natural Language Toolkit. In *Proceedings of the ACL-02 Workshop on Effective tools and methodologies for teaching natural language processing and computational linguistics - Volume 1* (pp. 63–70). Philadelphia, Pennsylvania: Association for Computational Linguistics.

Luwel, M. (2021). Performance-based Institutional Research Funding in Flanders, Belgium. *Scholarly Assessment Reports, 3*(1). doi:https://doi.org/10.29024/sar.29

Macaluso, B., Pollitt, A., Gunashekar, S., & Larivière, V. (2015). *Comparative Scientometric Assessment of the Results of ERC Funded Projects: Bibliometric assessment report (D5)*.

Manathunga, C., & Brew, A. (2012). Beyond Tribes and Territories: New Metaphors for New Times: . In P. Trowler, M. Saunders, & V. Bamber (Eds.), *Tribes and Territories in the 21st Century* (pp. 58-70). New York: Routledge.

Marres, N. & de Rijcke, S. (2020). From indicators to indicating interdisciplinarity: A participatory mapping methodology for research communities in-the-making. Quantitative Science Studies, 1(3), 1041-1055. doi:https://doi.org/10.1162/qss_a_00062

Massey, D. (1999). Negotiating Disciplinary Boundaries. *Current Sociology, 47*(4), 5-12. doi:https://doi.org/10.1177/0011392199047004003

Matwin, S., & Sazonova, V. (2012). Direct comparison between support vector machine and multinomial naive Bayes algorithms for medical abstract classification. *Journal of the American Medical Informatics Association, 19*(5), 917. doi:https://doi.org/10.1136/amiajnl-2012-001072

McCain, K. W. (1998). Neural networks research in context: A longitudinal journal cocitation analysis of an emerging interdisciplinary field. *Scientometrics, 41*, 389-410. doi:https://doi.org/10.1007/BF02459053

McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426v3*.

Medelyan, O., & Witten, I. H. (2006). Measuring inter-indexer consistency using a thesaurus. In *6th ACM/IEEE-CS joint conference on Digital libraries* (pp. 296-297). Chapel Hill, NC, USA: ACM.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems* (Vol. 2, pp. 3111-3119).

Morris, S. A., & Van der Veer Martens, B. (2009). Mapping research specialties. *Annual Review of Information Science and Technology, 42*(1), 213-295. doi:https://doi.org/10.1002/aris.2008.1440420113

Moschitti, A., & Basili, R. (2004). Complex Linguistic Features for Text Classification: A Comprehensive Study. In S. McDonald & J. Tait (Eds.), *Advances in Information Retrieval. ECIR 2004.* (Vol. 2997, pp. 181-196). Sunderland, UK: Springer.

Mullins, N. C., Hargens, L. L., Hecht, P. K., & Kick, E. (1977). The Group Structure of Cocitation Clusters: A Comparative Study. American Sociological Review, 42(4), 552-562. doi:https://doi.org/2094554

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the Social Sciences and Humanities: A review. *Scientometrics, 66*(1), 81-100. doi:https://doi.org/10.1007/s11192-006-0007-2

Nichols, L. G. (2014). A topic model approach to measuring interdisciplinarity at the National Science Foundation. *Scientometrics, 100*, 741-754. doi:https://doi.org/10.1007/s11192-014-1319-2

Noichl, M. (2019). Modeling the structure of recent philosophy. *Synthese, 198*, 5089-5100. doi:https://doi.org/10.1007/s11229-019-02390-8

Norris, M., & Oppenheim, C. (2007). Comparing alternatives to the Web of Science for coverage of the social sciences' literature. *Journal of Informetrics, 1*(2), 161-169. doi:https://doi.org/10.1016/j.joi.2006.12.001

Núñez, R., Allen, M., Gao, R., Rigoli, C. M., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behaviour, 3*(8), 782-791. doi:https://doi.org/10.1038/s41562-019-0626-2

OECD. (2007). *Revised Fields of Science and Technology (FOS) Classification in the Frascati Manual*. Paris: OECD Publishing

Ossenblok, T., Engels, T. C. E., & Sivertsen, G. (2012). The representation of the social sciences and humanities in the Web of Science—a comparison of publication patterns and incentive structures in Flanders and Norway (2005–9). *Research Evaluation, 21*(4), 280-290. doi:https://doi.org/10.1093/reseval/rvs019

Ostrom, E. (2007). Challenges and growth: the development of the interdisciplinary field of institutional analysis. *Journal of Institutional Economics, 3*(3), 239-264. doi: https://doi.org/10.1017/S1744137407000719

Pauwels, L., & Verhage, A. (2019). Criminology in Belgium: from embryonic conception to contemporary currents in a nutshell: some food for thought. *Criminology in Europe, 18*(2), 9-20.

Pedersen, D. B. (2016). Integrating social sciences and humanities in interdisciplinary research. *Palgrave Communications, 2*(1). doi: https://doi.org/10.1057/palcomms.2016.36

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, É. (2011). Scikit-learn: Machine

Learning in Python. *Journal of Machine Learning Research, 12*(85), 2825-2830.

Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics, 81*(3), 719-745. doi: https://doi.org/10.1007/s11192-008-2197-2

Porter, A. L., & Rossini, F. A. (1985). Peer review of interdisciplinary research proposals. *Science, Technology, & Human Values, 10*(3), 33-38.

Pudovkin, A. I., & Garfield, E. (2002). Algorithmic procedure for finding semantically related journals. *Journal of the American Society for Information Science and Technology, 53*, 1113-1119.

Rafols, I. (2014). Knowledge Integration and Diffusion: Measures and Mapping of Diversity and Coherence. In *Measuring Scholarly Impact* (pp. 169-190).

Rafols, I. (2019). S&T indicators in the wild: Contextualization and participation for responsible metrics. *Research Evaluation, 28*(1), 7-22. doi:https://doi.org/10.1093/reseval/rvy030

Rafols, I. (2021). 'Measuring' interdisciplinarity: from indicators to indicating. Retrieved from:https://i2insights.org/2021/02/09/measuring-interdisciplinarity/

Rafols, I., & Meyer, M. (2009). Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics, 82*(2), 263-287. doi:https://doi.org/10.1007/s11192-009-0041-y

Read, J. (2010). *Scalable Multi-label Classification.* (Doctor of Philosophy (PhD) Doctoral thesis). University of Waikato, Hamilton, New Zealand. Retrieved from https://hdl.handle.net/10289/4645

Read, J., Pfahringer, B., Holmes, G., & Frank, E. (2009). Classifier Chains for Multi-label Classification. In W. Buntine, M. Grobelnik, D. Mladenić, & J. Shawe-Taylor (Eds.), *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2009.* Bled, Slovenia: Springer.

Rennie, J. D. M., Shih, L., Teevan, J., & Karger, D. R. (2003). Tackling the Poor Assumptions of Naïve Bayes Text Classifiers. In T. Fawcett & N. Mishra (Eds.), *Twentieth International Conferene on Mahine Learning (ICML-2003)* (pp. 616-623). Washington D.C., USA: The AAAI Press.

Richelle, M., Janssen, P., & Bredart, S. (1992). Psychology in Belgium. *Annual Review of Psychology, 43*(1), 505-529. doi: https://doi.org/10.1146/annurev.ps.43.020192.002445

Rinia, E. J., van Leeuwen, T. N., Bruins, E. E. W., van Vuren, H. G., & van Raan, A. F. J. (2002). Measuring knowledge transfer between fields of science. *Scientometrics, 54*(3), 347-362. doi: https://doi.org/10.1023/A:1016078331752

Rip, A., & Courtial, J.-P. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics, 6*(6), 381-400. doi:https://doi.org/10.1007/BF02025827

Rousseau, R., Zhang, L., & Hu, X. (2019). Knowledge Integration: Its Meaning and Measurement. In *Springer Handbook of Science and Technology Indicators* (pp. 69-94).

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53-65.

Rylance, R. (2015). Grant giving: Global funders to focus on interdisciplinarity. *Nature, 525*(7569), 313-315. doi: https://doi.org/10.1038/525313a

Schoepflin, U., & Glänzel, W. (2001). Two decades of "Scientometrics": An interdisciplinary field represented by its leading journal. *Scientometrics, 50*(2), 301-312.

Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics, 59*(3), 425-465. doi:https://doi.org/10.1023/B:SCIE.0000018542.71314.38

National Academy of Sciences (2004). *Facilitating Interdisciplinary Research*.

Sculley, D. (2010). Web-scale k-means clustering. In *WWW '10: Proceedings of the 19th International Conference on World Wide Web* (pp. 1177-1178).

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computer Surveys, 34*(1), 1-47. doi:https://doi.org/10.1145/505282.505283

Sechidis, K., Tsoumakas, G., & Vlahavas, I. (2011). On the Stratification of Multi-Label Data. In D. Gunopulos, T. Hofmann, D. Malerba, & M. Vazirgiannis (Eds.), *Machine Learning and Knowledge Discovery in Databases. Joint European Conference on Machine Learning and Knowledge Discovery in Databases - ECML PKDD 2011* (Vol. 6913). Athens, Greece: Springer.

Seeber, M., Vlegels, J., & Cattaneo, M. (2022). Conditions that do or do not disadvantage interdisciplinary research proposals in project evaluation. *Journal of the Association for Information Science and Technology*. doi:https://doi.org/10.1002/asi.24617

Shneider, A. M. (2009). Four stages of a scientific discipline; four types of scientist. *Trends in Biochemical Sciences, 34*(5), 217-223. doi:https://doi.org/10.1016/j.tibs.2009.02.002

Sievert, M. C., & Andrews, M. J. (1991). Indexing consistency in Information Science Abstracts. *Journal of the American Society for Information Science, 42*(1), 1-6. doi: https://doi.org/10.1002/%28SICI%291097-4571%28199101%2942%3A1<1%3A%3AAID-ASI1>3.0.CO%3B2-9

Sīle, L., Guns, R., Vandermoere, F., Sivertsen, G., & Engels, T. C. E. (2021). Tracing the context in disciplinary classifications: A bibliometric pairwise comparison of five classifications of journals in the social sciences and humanities *Quantitative Science Studies, 2*(1), 65-88. doi:https://doi.org/10.1162/qss_a_00110

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. doi:https://doi.org/10.1016/j.joi.2017.12.006

Sjögårde, P., & Ahlgren, P. (2019). Granularity of algorithmically constructed publication-level classifications of research publications: Identification of specialties. *Quantitative Studies of Science, 1*. doi:https://doi.org/10.1162/qss_a_00004

Small, H., & Crane, D. (1979). Specialties and disciplines in science and social science: An examination of their structure using citation indexes. *Scientometrics, 1*, 445-461.

Sorower, M. S. (2010). *A literature survey on algorithms for multi-label learning*.

Stichweh, R. (1992). The Sociology of Scientific Disciplines: On the Genesis and Stability of the Disciplinary Structure of Modern Science. *Science in Context, 5*(1), 3-15. doi: https://doi.org/10.1017/S0269889700001071

Stichweh, R. (2003). Differentiation in science: Causes and consequences. In G. H. Hardon (Ed.), *Unity of Knowledge in Transdisciplinary Research for Sustainable Development* (Vol. 1, pp. 82-90). Oxford, United Kingdom: EOLSS Publishers Co. Ltd.

Sugimoto, C. R., & Weingart, S. (2015). The kaleidoscope of disciplinarity. *Journal of Documentation, 77*(4), 775-794. doi:https://doi.org/10.1108/JD-06-2014-0082

Suominen, A., & Toivanen, H. (2016). Map of science with topic modeling: Comparison of unsupervised learning and human-assigned subject classification. *Journal of the Association for Information Science and Technology, 67*(10), 2464-2476. doi:https://doi.org/10.1002/asi.23596

Suominen, O. (2019). Annif: DIY automated subject indexing using multiple algorithms. *LIBER Quarterly, 29*(1), 1-25. doi:http://doi.org/10.18352/lq.10285

Tarrant, S. P., & Thiele, L. P. (2017). Enhancing and promoting interdisciplinarity in higher education. *Journal of Environmental Studies and Sciences, 7*, 355-360. doi:https://doi.org/10.1007/s13412-016-0402-9

Thijs, B. (2020). Using neural-network based paragraph embeddings for the calculation of within and between document similarities. *Scientometrics, 125*, 835-849. doi:https://doi.org/10.1007/s11192-020-03583-6

TIjssen, R. J. W. (1992). A quantitative assessment of interdisciplinary structures in science and technology: Co-classification analysis of energy resarch. *Research Policy, 21*(1), 27-44. doi:https://doi.org/10.1016/0048-7333(92)90025-Y

Tomov, D., & Mutafov, H. (1996). Comparative indicators of interdisciplinarity in modern science. *Scientometrics, 37*(2), 267-278. doi:https://doi.org/10.1007/bf02093624

Truc, A., Santerre, O., Gingras, Y., & Claveau, F. (2020). The Interdisciplinarity of Economics. *SSRN Scholarly Paper ID 3669335, Rochester, NY: Social Science Research Network.*

Tsoumakas, G., & Katakis, I. (2007). Multi-Label Classification. *International Journal of Data Warehousing and Mining, 3*(3), 1-13. doi:https://doi.org/10.4018/jdwm.2007070101

Turner, S. (2000). 3. What Are Disciplines? And How Is Interdisciplinarity Different? In *Practising Interdisciplinarity* (pp. 46-65).

Urata, H. (1990). Information flows among academic disciplines in Japan. *Scientometrics, 18*(3), 309-319. doi:https://doi.org/10.1007/BF02017767

Urban, D. (1982). Mobility and the Growth of Science. *Social Studies of Science, 12*(3), 409-433. Retrieved from http://www.jstor.org/stable/284667

van Baalen, P., & Karsten, L. (2012). The evolution of management as an interdisciplinary field. *Journal of Management History, 18*(2), 219-237. doi:https://doi.org/10.1108/17511341211206861

van den Besselaar, P. (2019a, 11th April 2019). *Interdisciplinary and disciplinary identities: towards a theory of forms of knowledge change*. pre-print. BioRxiv.

van den Besselaar, P. (2019b). Interdisciplinary and disciplinary identities: towards a theory of forms of knowledge change. *bioRxiv*, 603449. doi:https://doi.org/10.1101/603449 van den Besselaar, P., & Heimeriks, G. (2001). *Disciplinary, multidisciplinary, interdisciplinary: Concepts and indicators.* Paper presented at the International Conference on Scientometrics & Informetrics, Sydney, Australia.

van den Besselaar, P., & Heimeriks, G. (2006). Mapping research topics using word-reference co-occurrences: A method and an exploratory case study. *Scientometrics, 63*(3), 377-393. Retrieved from https://akademiai.com/doi/abs/10.1007/s11192-006-0118-9

van den Besselaar, P., & Leydesdorff, L. (1996). Mapping change in scientific specialties: A scientometric reconstruction of the development of articial intelligence. *Journal of the American Society for Information Science, 47*(6), 415-436. doi: https://doi.org/10.1002/(SICI)1097-4571(199606)47:6<415::AID-ASI3>3.0.CO;2-Y

van der Maaten, L., & Hinton, G. (2008). Visualising Data using t-SNE. *Journal of Machine Learning Research, 9*(86), 2579-2605.

van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523-538. doi:https://doi.org/10.1007/s11192-009-0146-3

van Houten, J., van Vuren, H. G., Le Pairs, C., & Dijkhuis, G. (1983). Migration of physicists to other academic disciplines: Situation in the Netherlands. *Scientometrics, 5*(4), 257-267. doi:https://doi.org/10.1007/BF02019741

Van Praag, L., & Daenekindt, S. (2021). Beyond Disciplinary Blind Spots: A Systematic Review of Research on Ethnicity and Race in Secondary Education Using Automated Text Analysis. *New Diversities, 23*(1), 105-121.

Vancauwenbergh, S., & Poelmans, H. (2019a). The creation of the Flemish research discipline list, an important step forward in harmonising research information (systems). *Procedia Computer Science, 146*, 265-278. doi:https://doi.org/10.1016/j.procs.2019.01.075

Vancauwenbergh, S., & Poelmans, H. (2019b). The Flemish Research Discipline Classification Standard: A Practical Approach.

Knowledge Organisation, 46, 354-363. doi:https://doi.org/10.5771/0943-7444-2019-5-354

Vandermoere, F., & Vanderstraeten, R. (2012). Disciplinary networks and bounding: Scientific communication between Science and Technology Studies and History of Science. *Minerva, 50*(4), 451-470. doi:https://doi.org/10.1007/s11024-012-9210-x

Vanderstraeten, R. (2021). Publikations- und Evaluationspraktiken in wissenschaftlichen Zeitschriften, 1930–2020: Eine Fallstudie. *Zeitschrift für Soziologie, 50*(3-4), 274-287. doi:https://doi.org/10.1515/zfsoz-2021-0019

Vanderstraeten, R., & Vandermoere, F. (2015). Disciplined by the Discipline: A Social-Epistemic Fingerprint of the History of Science. *Science in Context, 28*(2), 195-214. doi:https://doi.org/10.1017/s0269889715000058

Vanderstraeten, R., & Vandermoere, F. (2021). Inequalities in the growth of Web of Science. *Scientometrics, 126*(10), 8635-8651. doi:https://doi.org/10.1007/s11192-021-04143-2

Verleysen, F., Ghesquière, P., & Engels, T. (2014). The objectives, design and selection process of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW). In W. Blockmans, L. Engwall, & D. Weaire (Eds.), *Bibliometrics: Use and abuse in the review of research performance* (Vol. 87, pp. 117-127). London: Portland Press Ltd.

Viard, A., Eustache, F., & Segobin, S. (2021). History of Magnetic Resonance Imaging: A Trip Down Memory Lane. *Neuroscience, 474*, 3-13. doi:https://doi.org/10.1016/j.neuroscience.2021.06.038

Viseu, A. (2015). Integration of social science into research is crucial. *Nature, 525*(7569), 291-291. doi:https://doi.org/10.1038/525291a

Vugteveen, P., Lenders, R., & Van den Besselaar, P. (2014). The dynamics of interdisciplinary research fields: the case of river research. *Scientometrics, 100*(1), 73-96. doi:https://doi.org/10.1007/s11192-014-1286-7

Wagner, C. S., Roessner, D. J., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., . . . Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics, 5*(1), 14-26. doi:https://doi.org/10.1016/j.joi.2010.06.004

Waltman, L., Boyack, K. W., Colavazzi, G., & Van Eck, N. J. (2020). A principled methodology for comparing relatedness measures for clustering publications. *Quantitative Science Studies, 1*(2). doi: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. doi:https://doi.org/10.1002/asi.22748

Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and Impact: Distinct Effects of Variety, Balance, and Disparity. *PLOS ONE, 10*(5). doi:https://doi.org/10.1371/journal.pone.0127298

Wang, Q., & Schneider, J. W. (2019). Consistency and validity of interdisciplinarity measures. *Quantitative Studies of Science*, 1-25. doi:https://doi.org/10.1162/qss_a_00011

Wang, S., & Koopman, R. (2017). Clustering articles based on semantic similarity. *Scientometrics, 111*(2), 1017-1031. doi:https://doi.org/10.1007/s11192-017-2298-x

Waskom, M. L. (2021). Seaborn: statistical data visualization. *Journal of Open Source Software, 6*(60), 3021. doi:https://doi.org/10.21105/joss.03021

Weingart, P. (2003). Growth, Differentiation, Expansion and Change of Identity — The Future of Science. In *Social Studies of Science and Technology: Looking Back, Ahead* (pp. 183-200).

Weingart, P. (2010). A short history of knowledge formations. In J. T. Klein & C. Mitcham (Eds.), *The Oxford Handbook of Interdisciplinarity* (pp. 3-14). Oxford: Oxford University Press.

Weingart, P. (2016). From “Finalization” to “Mode 2”: old wine in new bottles? *Social Science Information, 36*(4), 591-613. doi:https://doi.org/10.1177/053901897036004002

Whitley, R. (2000). *The intellectual and social organization of the sciences* (2nd ed. ed.). Oxford: Oxford University Press.

Wright, B. E. (2011). Public Administration as an Interdisciplinary Field: Assessing its Relationship with the Fields of Law, Management, and Political Science. *Public Administration Review, 71*(1), 96-101. doi:https://doi.org/10.1111/j.1540-6210.2010.02310.x

Wyatt, S. (2016). A computational turn in the humanities? a perspective from science and technology studies. *Журнал Сибирского федерального университета. Гуманитарные науки, 9*(2), 517-524

Yan, E. (2014). Finding knowledge paths among scientific disciplines. *Journal of the Association for Information Science and Technology, 65*(11), 2331-2347. doi:https://doi.org/10.1002/asi.23106

Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other. *Journal of the American Society for Information Science and Technology, 63*(7), 1313-1326. doi:https://doi.org/10.1002/asi.22680

Yan, E., Ding, Y., Cronin, B., & Leydesdorff, L. (2013). A bird's-eye view of scientific trading: Dependency relations among fields of science. *Journal of Informetrics, 7*(2), 249-264. doi:https://doi.org/10.1016/j.joi.2012.11.008

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. doi:https://doi.org/10.1016/j.joi.2011.10.001

Yau, C.-K. P., Alan; Newman, Nils; Suominen, Arho. (2014). Clustering scientific documents with topic modeling. *Scientometrics, 100*, 767-786. doi:https://doi.org/10.1007/s11192-014-1321-8

Yegros-Yegros, A., Rafols, I., & D'Este, P. (2015). Does Interdisciplinary Research Lead to Higher Citation Impact? The Different Effect of Proximal and Distal Interdisciplinarity. *PLOS ONE, 10*(8). doi:https://doi.org/10.1371/journal.pone.0135095

Yeo, R. (1991). Reading Encyclopedias: Science and the Organization of Knowledge in British Dictionaries of Arts and Sciences, 1730-1850. *Isis, 82*(1), 24-49. doi:https://doi.org/10.1086/355636

Yu, X., Szymanski, B. K., & Jia, T. (2021). Become a better you: Correlation between the change of research direction and the change of scientific performance. *Journal of Informetrics, 15*(3), 101193. doi:https://doi.org/10.1016/j.joi.2021.101193

Zhang, L., Rousseau, R., & Glänzel, W. (2015). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *JASIST - Journal of the Association for Information Science and Technology, 67*(5), 1257-1265. doi:https://doi.org/10.1002/asi.23487

Zhang, M.-L., Li, Y.-K., Liu, X.-Y., & Geng, X. (2018). Binary relevance for multi-label learning: an overview. *Frontiers of Computer Science, 12*, 191-202. doi:https://doi.org/10.1007/s11704-017-7031-7

Zhang, M.-L., & Zhou, Z.-H. (2014). A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering, 26*(8), 1819-1837. doi:https://doi.org/10.1109/tkde.2013.39

Zhou, H., Guns, R., & Engels, T. C. (2021). The evolution of interdisciplinarity in five social sciences and humanities disciplines: relations to impact and disruptiveness. In W. Glänzel, S. Heeffer, P.-S. Chi, & R. Rousseau (Eds.), *Proceedings of the 18th conference of the International Society for Scientometrics and Informetrics (ISSI 2021)* (pp. 1381-1392). Belgium: ISSI.