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

Joshua Eykens¹ Raf Guns² Raf Vanderstraeten³

¹ Centre for R&D Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Antwerp (Belgium); joshua.eykens@uantwerpen.be; ORCID: 0000-0002-1680-0112

² Centre for R&D Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Antwerp (Belgium); raf.guns@uantwerpen.be; ORCID: 0000-0003-3129-0330

³ Department of Sociology, Ghent University, Ghent (Belgium) and Department of Sociology, The London School of Economics and Political Science, London (GB); raf.vanderstraeten@ugent.be; ORCID: 0000-0002-4155-2499

ABSTRACT

In this study we explore the disciplinary diversity present within subject specialties in the social sciences and humanities. Subject specialties are operationalized as textually coherent clusters of documents. We apply topic modelling to textual information on the individual document level (titles and abstracts) to cluster a multilingual set of roughly 45,000 documents into subject specialties. The dataset includes the metadata of journal articles, conference proceedings, book chapters, and monographs. We make use of two indicators, namely, the organizational affiliation based on the departmental address of the authors and the cognitive orientation based on the disciplinary classifications at the publication level. First, we study the disciplinary diversity of the clusters by calculating a Hill-type diversity index. We draw an overall picture of the distribution of subject specialties over diversity scores and contrast the two indicators with each other. The goal is to discover whether some subject specialties are inherently multi- or interdisciplinary in nature, and whether the different indicators are telling a well-aligned, similar story. Second, for each cluster of documents we calculate the dominance, i.e. the relative size of the largest discipline. This proxy of disciplinary concentration gives an idea of the extent to which a specialty is disciplined. The results show that all subject specialties analyzed serve as interdisciplinary trading grounds, with outliers in both directions of the disciplinary-interdisciplinary continuum. For a large share of specialties, the dominant cognitive and organizational disciplinary classification were found to be well aligned. We present a typology of subject specialties by contrasting the organizational and cognitive diversity scores.

Keywords: *document clustering, interdisciplinarity, social sciences and humanities, specialization, textual data, Top2Vec*

1. Introduction

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 labor 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.

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 analyze 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) 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). 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 neighboring 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, Santerre, Gingras, and Claveau (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 analyze the disciplinary and interdisciplinary diversity of science. We particularly aim at developing a bottom-up approach to analyze 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 (Hammarfelt, 2018; Sugimoto & Weingart, 2015). This has resulted in numerous studies investigating the usefulness of different methods to operationalize the concept (e.g., Guns, Sīle, Eykens, Verleysen, & Engels, 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 silo’s.

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 (edition X), 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, Ghesquière, & Engels, 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 refined OECD FOS classification system, for details see: Guns et al., 2018). and one that is based on the organizational affiliation of the authors (Guns et al., 2018).

Table 1. Number of records per publication type and most prominent language category in the final dataset used for further analyses.

	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

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.

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, Corrado, & Dean, 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).

3.1 Top2Vec

As an unsupervised topic modeling or clustering solution, Top2Vec (Angelov, 2020) has two main benefits: (1) it is entirely unsupervised, no preset 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).

3.1.1 Document vectorization: Universal Sentence Encoders

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.

3.1.2 Document clustering: UMAP + HDBSCAN

After embedding the documents and words, Uniform Manifold Approximation and Projection or UMAP (McInnes, Healy, & Melville, 2018) 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.

Hierarchical DBSCAN (HDBSCAN) (Campello, Moulavi & Sander, 2013), is applied to the UMAP projection to find dense areas of similar documents. 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.

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.

3.2.1 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 $p_i = x_i / \sum_{k=1}^n x_k$.

3.2.2 (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 (Zhang, Rousseau & Glänzel, 2015; Jost, 2009) 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.

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.

4.1 Clustering outcomes

Figure 1 shows a 2D UMAP projection of the clusters. The largest clusters are labeled 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.

Figure 1. UMAP projection of the document vectors. The 11 largest clusters are highlighted in color and labeled with the first five topic words (according to Top2Vec).

Figure 2. Distribution of clusters over different cluster sizes.

4.2 The disciplinary diversity of subject specialties

From figure 3 it becomes clear that the diversity found for the subject specialties varies a lot. For both classification systems the center 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.

Figure 3. Density distribution of disciplinary diversity for all subject specialties. Based on the organizational affiliation of authors and the cognitive disciplinary classification.

We now turn to the question how disciplinary diversity on the input side is related to disciplinary diversity on the output side. Figure 4 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 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.

Figure 4. 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. The labels of the dots correspond to the index number of the subject specialties. They are shown on the right of the dots or have been connected by a line.

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.

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.

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 5.

Figure 5. 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 .40 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 4) 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 modeling, 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 5), these examples illustrate that they are not

interchangeable and tell a different story about either the specialization or (cross-)disciplinarity of a subject specialty.

4.3 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 6. 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 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 is 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 7 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.

Figure 7. 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.

4.4 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 8. 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.

Whereas it was indicated above that economics and business is less cross-disciplinary than their neighboring 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.

1. 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-exists 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

Figure 8. 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.

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 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, Santerre, Gingras, & Claveau, 2020), while sociology has often been quoted as a very ‘open’ discipline, sharing many subjects with other disciplines (Abbott, 2010; 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.

5.1 Limitations

The Top2Vec algorithm used to delineate subjects 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 surplus 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.

The use of a local database imposes some limitations. 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.

5.2 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?

2. 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.

Bibliography

- Abbott, A. D. (2010). *Chaos of Disciplines*. Chicago: University of Chicago Press.
- Angelov, D. (2020). Top2Vec: Distributed Representations of Topics. *Arxiv preprint arXiv:2008.09470*.
- Campbell, D. T. (1969). Ethnocentrism of disciplines and the fish-scale model of omniscience. In M. Sherif & C. W. Sherif (Eds.), *Interdisciplinary Relationships in the Social Sciences* (pp. 328-348). Chicago: Aldine.
- Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, & G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science* (Vol. 7819, pp. 160-172). Berlin, Heidelberg, Germany: Springer.
- Cer, D., Yang, Y., Kong, S.-Y., Hua, N., Limtiaco, N., John, R. S., et al. (2018). Universal Sentence Encoder. *arXiv preprint ArXiv:1803.11175v2*.
- Durkheim, E. (2014). *The division of labor in society*. New York, N.Y.: Free Press.
- Engels, T. C. E., & Guns, R. (2018). 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>
- 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.
- 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.
- Graff, H. J. (2015). *Undisciplining Knowledge: Interdisciplinarity in the Twentieth Century*. Baltimore, Maryland: Johns Hopkins University Press.
- Guns, R., Sile, 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>

- 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.
- 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>
- Jacobs, J. (2013). *In Defense of Disciplines: Interdisciplinarity and Specialization in the Research University*. Chicago: University of Chicago Press.
- 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>
- 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>
- 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.
- 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>

- 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. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv preprint*. arXiv:1802.03426
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint*. arXiv:1301.3781.
- Noichl, M. (2019). Modeling the structure of recent philosophy. *Synthese*, 198, 5089-5100. doi: <https://doi.org/10.1007/s11229-019-02390-8>
- 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>
- 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>
- 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>
- Schoepflin, U., & Glänzel, W. (2001). Two decades of "Scientometrics": An interdisciplinary field represented by its leading journal. *Scientometrics*, 50(2), 301-312. doi: <https://doi.org/10.1023/A:1010577824449>
- Sîle, 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
- 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. doi: <https://doi.org/10.1007/BF02016661>
- 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>
- Truc, A., Santerre, O., Gingras, Y., & Claveau, F. (2020). *The Interdisciplinarity of Economics*. SSRN Scholarly Paper ID 3669335, Rochester, NY: Social Science Research Network.

- 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. (2019, 11th April 2019). *Interdisciplinary and disciplinary identities: towards a theory of forms of knowledge change*. pre-print. *BioRxiv*.
- Van den Besselaar, P., & Leydesdorff, L. (1996). Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial 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](https://doi.org/10.1002/(SICI)1097-4571(199606)47:6<415::AID-ASI3>3.0.CO;2-Y)
- 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. & 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>
- Verleysen, F. T., Ghesquière, P., & Engels, T. C. E. (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* (pp. 117-127). London: Portland Press.
- Vugteveen, P., Lenders, R., & Van den Besselaar, P. (2014). The dynamics of interdisciplinary research fields: The case of river research. *Scientometrics*, 100, 73-96. doi: <https://doi.org/10.1007/s11192-014-1286-7>
- 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>
- 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>
- 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>