
ADULT EDUCATION CENTRES - PLACES OF COMPLEMENTARY VOCATIONAL EDUCATION AND TRAINING? A COMPARATIVE STUDY OF THE CONTENTS AND CONTEXTS


A WORKING PAPER

 **Tanja Preböck***

Chair of Human and Business Education
University of Bamberg
tanja.preboeck@uni-bamberg.de

 **Alexander Brand**

Department of Social Sciences
University of Hildesheim
alexander.brand@uni-hildesheim.de

 **Silvia Annen**

Chair of Human and Business Education
University of Bamberg
silvia.annen@uni-bamberg.de

September 16, 2021

ABSTRACT

Lifelong learning is a fundamental facet of the modern educational landscape, especially in the vocational context. Vocational competences need to be continuously developed, starting from different educational or qualification levels, for which extensive formal and non-formal educational offers exist. The adult education centres (vhs) are a relevant actor here. Their offerings in this area present planners with the macro-didactic challenge of making courses with very different content and varying levels attractive to a broad audience. This Article analyses the programmes of the Adult Education Centres (VHS / Volkshochschulen) in Germany, retrieve-able via their online Course Programme ($N = 53942$) using topic modelling and word embedding regression models, a closer look will be taken on to (sub)categories, topics and contexts of courses. Our results show an unequal distribution of course categories the prevalence of cross-sectional topics, like "Basic digital skills", "Emotional Well-Being", and "World Knowledge". While these topical and categorical sets of information show the connections in courses as a whole; contexts may differ. Differing between terms like "Bilanz" (balance sheet) and "Buchführung" (accounting) we are able to show that specific prerequisites and usage arise even within the same topic and sub-category, respectively. Conclusively we show that the immense data of course description of VHS in Germany can be categorised roughly by educational structures, content and diverse formulations of prerequisites.

Keywords Adult Education · Structural topic model · Word Embedding · Volkshochschulen

1 Introduction

Lifelong learning is a fundamental facet of the contemporary educational landscape, especially in the vocational and professional context [Gartenschlaeger, 2017]. This is rather important, because knowledge and competences become more and more important in our society through macro-trends like the ongoing globalisation and the general tendency of modern societies in relying on highly educated individuals for reproducing value and prosperity through employability [Singh and Ehlers, 2020, Hinzen, 2020, Arnold et al., 2018]. So there is no astonishment that policies supporting lifelong learning and educating people outside the educational system are gaining relevancy [Gartenschlaeger, 2017, der Europäischen Gemeinschaften, 2000] and employers as well as employees are aware of the need to "stay up to

*Corresponding author

date" [Boffo, 2020, Singh and Ehlers, 2020]. In the German educational system, a place dedicated to such processes is the VHS-system (Adult Education Centres / Volkshochschulen), an institution that takes that place because of its relatively broad focus [Avramovska et al., 2017]. Opposite to formally structured and federal organised parts of the educational system, like elementary and secondary school phase [Kotthoff, 2011], VHS, which can be referred to as quartary education (after universities being tertiary) are relatively free in their choice of courses (with the exception of courses offering more formal qualifications, like integration proofs), offering a wide variety of thematic variations [Süssmuth and Eisfeld, 2018]. While this open focus offers easy, beginner-friendly and more flexible learning structures, it comes with some problems: One specific problem is the insecurity of employers and employees looking at such institutions with specific further-education needs. Whilst first attempts have been made to address this issue, such as the Xpert Business for categorising courses related to further education, a certain degree of insecurity remains, as this classification is neither compulsory nor all-encompassing [Xpert Business, 2021]. We address this problem in our paper by proposing a first exploratory characterisation via means of quantitative text analysis. For this we aim to provide insights into understanding the structuring and usage of (sub)categories, topics and relevant terms. The focus will be on the semantic level in order to be able to analyse the external self-presentation of the Adult Education Centres and the content links of the course offerings. For this we look at existing categorisations, extract latent topics via topic modelling and look at prototypical contexts for some further education specific terms.

The article starts with a short overview over the state of research on the VHS, focusing its general structure as well as their role in further education, and its program planning. After discussing the state of research and defining relevant terms for the formal modelling, we discuss the used data, methodology and methods. The following results can be separated into descriptive (ratio of offered courses by category and subcategory, frequency of topics, and share of topics by category) and conclusive (regression model and near neighbours). Subsequently, the results are summarised and particularly interesting results are discussed. The article closes with a conclusion and then gives an outlook on limitations and further research approaches.

2 State of Research and Theoretical Foundations

Generally spoken, the educational institutions focused here, Adult Education Centres or VHS, are considered as "local learning places that are accessible for a broad sector of the population" [Avramovska et al., 2017, p. 5]. Contrary to formal education, VHS don't focus on a general age for their participants, most courses nevertheless address adults (some events explicitly seniors), while others contrarily exclusively focus children in special courses for them [Hinzen, 2020]. In the German educational landscape, education generally starts with kindergarten (3 to 6 years). Afterwards pupils visit an elementary school. Afterwards, an early separation (mostly after grade 4) of pupils into 2 to 3 tracks of comprehensiveness, with different duration (6 to 9 years) and achievable school-leaving certificates, occurs. Comprehensive schools are rare in secondary education and only present in some federal states. A rather specific issue is vocational training within the "dual system", which consists of an apprenticeship within a company and lessons in a vocational school. This normally takes three years. On the other hand vocational training colleges educate pupils themselves and organise practical experiences for their students around their schooling. Also a part of tertiary education are universities. Federal states are responsible for primary, secondary and tertiary education, nevertheless there is a body on the national level (Kultusministerkonferenz / KMK). As schools are regulated on a state level, municipalities can only slightly make an impact [Kotthoff, 2011].

World wide, VHS are mentioned as an example illustrating the anchoring of adult education in the broad society [Heuer and Kuprian, 2017]. The necessity for this is obvious, as, due to general demographic trends, we tend to be adults for the most part of our lives [Hinzen, 2020]. VHS, which would be referred to as quartary education, can be traced back to a more than 100 year old history in Germany [Gartenschlaeger, 2017]. As they are regionally organised (mostly on the municipal level) [Huntemann et al., 2019], unlike schools they can take regional characteristics into account, e. g. via teaching dialects or offering excursions in local history.

A great strength of AECs world wide is its adaptiveness. If a regional VHS is well-organised, it is expected that it can adapt easily to arising demands [Süssmuth and Eisfeld, 2018], such as, within the last years, the need for integration classes and the further arising demand for groups or courses for elder citizens [Gartenschlaeger, 2017]. Therefore, a given programme can always be viewed as a representation between societal needs in adult education and the specific position of an offering institution [Reich-Claassen and Hippel, 2018]. Accordingly, the programme offered by a VHS will be different from one offered by KAB (catholic workers movement) or a trade union.

As VHS are educational institutions, didactic is a relevant part of their daily business. It can be separated via its level into micro-, meso- and macro-didactic. Whilst micro-didactic refers to the individual teaching situation within a course; macro-didactic is the planning of offers and the programme and is considered to be the most important [Reich-Claassen and Hippel, 2018]. VHS usually provide a six month (semester) comprehensive programme, which consists of ongoing or repetitive courses and special offers. Throughout the week courses are offered on all days, including evenings and weekends, with specific age-groups preferring different times and days of the week [Süssmuth

and Eisfeld, 2018, Huntemann et al., 2019]. Analysing programmes in adult and further education is not new. Recent works used all manners of methods to focus on different aspects, taking into account various focus. Oftentimes they rely on the print-version of the programme and include aspects such as visualisation. Surprisingly, in the German-speaking countries there is no conservation archive of all AEC (or merely VHS)-programmes [Giseke et al., 2018]. Relevant questions while developing such a programme for one specific VHS (or VHS-association) are: Which course will I be offering when for whom? Meso-didactic, following this separation, would focus the planning of one course [Reich-Claassen and Hippel, 2018]. The difficulty within the field of programme planning is, that those in charge have to anticipate participants' needs and desires as well as to pull them into certain topics they consider relevant for their audience [Reich-Claassen and Hippel, 2018]. Therefore VHS function as a seismograph, as they illustrate social trends in their programme offers [Heuer and Kuprian, 2017]. It is a central part of the organisation of the course-structure of VHS, that they are demand-orientated and affordable and furthermore have no hurdles to overcome for participation [Gartenschlaeger, 2017], because VHS are only partially state-funded, the rest of their expenses must be covered by participants' fees [Heuer and Kuprian, 2017] (if further interested, see also: [Huntemann et al., 2019]). Therefore, highly demanded courses, such as Fitness and basic level foreign languages courses are offered at a lot of places a lot of different times throughout the week. During courses they describe themselves as "participant-centred" [Süssmuth and Eisfeld, 2018, 766], which implies that the content, teaching and structure of the course is defined through interaction with participants. On the other hand, (nearly) none of the courses has strict requirements of prior knowledge [Süssmuth and Eisfeld, 2018], but some of them recommend specific competencies to build up upon. If those are required, this explication often occurs in the "course description".

In the 1980s, due to the demand for higher performance on the job, EDP and health-courses, as well as a general shift towards coping with life, to the expense of educational knowledge, was and is still observable. Via offering the opportunity to acquire skills, know-how and personal capacities, the VHS are active in the field of employability [Boffo, 2020, Groth, 2018]. Further education, in addition to education along the job, become more relevant [Arnold et al., 2018]. Therefore cooperations between companies, as well as individual support (in cooperation with the Federal Employment Agency), and the VHS are also rising, as they are, additional to education near the job, regarded relevant by potential participants and employers [Arnold et al., 2018, Süssmuth and Eisfeld, 2018]. Accordingly, professional qualification and catch-up learning for employability are important within the programmes of VHS, in the field of Work - Job, they ideologically pursue a more holistic and integrative approach of learning and teaching enabling vocational and societal integration of persons, who (still) lack that (partially) [Süssmuth and Eisfeld, 2018]. Educational attainment in general shifts towards an "just in time"-Qualification, meaning that employers and employees seek for certain input, when a specific qualification is needed and do not anymore educate "on stock", where they would teach all knowledge during vocational training. A major reason seems to be the arise of the concept of a subsequent jobs, rather than a lifetime job, that is done from the end of training until retirement [Arnold et al., 2018].

A lot of courses, VHS offer in regard to Work - Job are offered to employees for coping with the changing requirements in their tasks due to elevation of technical support and work-place-organisation. Objectives hereby can be to impart special qualifications, deepen and revisit overarching fundamentals based on relevant problems and/or to take up new topics, to compensate for gaps in initial training, or to adapt qualifications to technological and economic innovations [Dittrich, 2018, Faulstich, 2018, ?]. Different customers hereby require different emphasis, due to their personal competencies and motivation, that drives them into further education. Through the combination of addressees, goals and contents, a general course-offer can be created. Adult Education Centres are gaining a more relevant role herein within the recent years [Faulstich, 2018]. Whilst a lot of big enterprises run their own further education department, especially employees from small and medium sized business take courses from external suppliers, and therefore can be considered VHS main target group [Faulstich, 2018].

After focusing on VHS in general and their role in further education, the participants of VHS-courses will be regarded. In 2019 most courses are offered in the category of Health (35 %), followed by Languages (32 %). Culture - Creative is the category for 16 % of courses, whilst Work - Job and Politics - Society - Environment are relatively egalitarian with 8 and 7 %, respectively. Basic education and School Certificates account for 1 % each [Huntemann et al., 2019]. According to the official VHS-Statistics from 2019 in the field of Work - Job, the majority of participants is female (65 %), as well as it is the majority of involved persons in all roles (participant, teacher and planner) [Süssmuth and Eisfeld, 2018, Huntemann et al., 2019]. Unsurprisingly in the category of Work - Job most participants are aged 50-65 (30 %) and 35-49 (25 %), respectively. Furthermore 13 % are aged 65-74 and 12 % 25-34. Only small shares of participants are under 18 (8 %), 75 and older (7 %) or 18-24 years old (6 %) [Huntemann et al., 2019]. That same statistics does not analyse course participants in regard to their formal educational level [Huntemann et al., 2019]. The general participant attending a course in the field of Work - Job accordingly is female and of working age, which can lead to the conclusion, that these courses can and might be intended as update and repetition of job-related competencies, maybe even after being off the job for a while (e.g. childcare break).

Although, as mentioned above, several aspects of Adult Education or VHS have already been analysed, a country wide analysis of the (sub-)categories, topics and contexts of the VHS course description data has not been conducted so far. This differentiation and practice seems to suit best here as it unifies three different layers of structure of courses.

The first layer is the (sub-)category. The VHS offer their courses in six, mutually exclusive, categories with each integrating several, amongst themselves again mutually exclusive, subcategories. Each category consists of a self-perception and fields of action. Such offer an insight in the organisational validation of the engroupings [Deutscher Volkshochschul-Verband e.V., 2021]. It is eye-catching, that categories, if descriptions are focused, are not mutually exclusive anymore, e. g. a specific course of "Business English" could fit into the categories of "Languages" or "Work - Job". Where it can be found nowadays is the result of mutual consent and might be subject to a future change through the DVV (German VHS Association). By offering a first scheme of categorisation those categories allow a rough differentiation of courses aligned with that and exclusively that. Unlike previous publications, here we take into account live data from 2020 to analyse the categorical scheme. This allows us to be up to date and to account for major changes that were made due to the COVID-19 pandemic and its restrictions.

A topic model version of the corpus is the second layer. Nearly all courses in the VHS-Kursfinder include a course description. This description includes further information, such as scope of the course, bring-alongs, etc. Such topical description offer deeper insights into the un-pronounced, latent requirements and framings of content information. This information are not accessible through the categories, as they are not compulsorily driven by course content. Furthermore these information are not pre-defined and therefore planned course contents can be adequately transmitted in order to generate a mutual understanding of it. On an analytical point of view this allows to an exploration of latent semantics and a rough combination of concepts, that are not joinable in the categories (e. g. in the Business-English course the concept of language skills and job demands).

The contextual layer dives even deeper as a third layer of analysis. Unlike the two others, it focuses on the word level and different practises of using words. This allows to further differentiate latent units of meanings in the sub-categories or topics. An analysis of this words and their co-occurrences accounts for linguistically different using of words in different contexts (e. g. in the Business-English course it is point out, that role-playing refers to a micro-didactic practice and not a creative way of expressing oneself). For identifying relevant terms in the field of business education, a special vocabulary was created. Therefore from the standard thesaurus economics (STW) the subthesauri V (Economics) and B (Business administration) were used and words reduced to their stems. The other subthesauri (Economic sectors, Products, Neighbouring sciences, Geographical terms and General terms) were excluded, mainly due to their generality as the purpose of extracting terms, that are relevant in the field of business would not be achievable including them. The choice for this thesaurus was made, because it is openly available, bilingual and was lately updated [ZBW, 2021]. Our final vocabulary ($V_{Business}$) contains the following 25 words: Balance sheet, Accounting, Management, Communication, Training (stem), Working hours, Product (stem), Economic (stem), Costs, Controlling, Work, Career, Employment, Personal, Computer, Capital, Human resources, Profit, Leadership, Enterprise, Performance, employees, Telecommunications, Data processing and Data.

Our proceeding in this paper combines the three presented layers to take an holistic approach towards courses offered in the VHS. In the following the methods and used data will be described.

3 Methods and Data

Methodologically, we applied a three step procedure, consisting of topic modelling (1), word embedding (2), and analysing the relationship between categories, topics and concepts via à la carte text embedding (regression models) (3) (with the substeps Parameter Estimation and Term Differentiation), to evaluate each previously defined textual level in our data. An visual overview of the methodological steps can be obtained in Figure 1. In the following steps we describe our used dataset, explain the ratios behind each methodological step very briefly and show the necessary substeps in each part of our analysis.

3.1 Data

For our analysis we collected textual (title, formal description of the offer) and categorical data (category, subcategory) of all course offers publicly available on the "VHS-Kursfinder" (retrievable via <https://www.volkshochschule.de/kursfinder>) on the 8th of December of 2020 via screen scraping. Due to some issues with the consistency of sorting into the specific fields in the formulas, we applied a manual check whether scraped information are correctly assigned to the specific columns, sorted if elsewhere and removed empty containers. Another necessary cleaning step was the removal of multilingual and non-german course descriptions to enable text processing via (structural) topic modelling and word embedding without language-wise alignment, which would be difficult due to very low sample sizes for non-german course descriptions in our dataset. Following this process we ended with a dataset consisting of $N = 53942$ course offers for our further analytical steps.

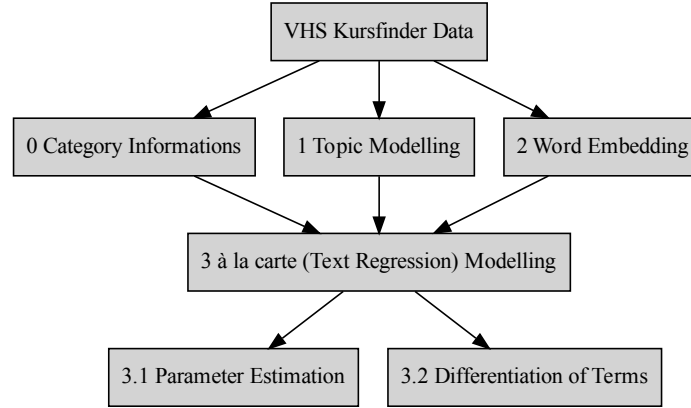


Figure 1: Analytical steps in the evaluation of the textual components of the course description in the VHS dataset

3.2 Topic Modelling

A topic model can be described as a language model with the task to annotate a collection of texts (in the following referred as the text corpora) with a previous unknown set of labels (= topics) [Fischer-Preßler et al., 2019]. This is based on the assumption that texts are a composition of latent topics, which can be described by groups of words encoding (humanly) interpretable themes in the documents of the text corpus [Blei et al., 2003]. Following this, we can fuzzy categorise documents by their composition of such themes (e.g. a course offer describeable by 90 % via one theme and 10 % by an other). While human coders with cross-coder reliability checks would be ideal for this type of task [Fischer-Preßler et al., 2019], the large amount of text in our corpus makes this somewhat difficult and enables us to use an automated content analysis tool widely applied in political and social science [Bail et al., 2017, Nelson, 2020, Schwemmer, 2021].

To preprocess our data for the topic modelling, we applied a stopwords removal for german terms on our bag-of-words approach (words are represented as features of a text, ignoring their order) and discarded irrelevant information like URLs, punctuation and numbers in the tokenisation step. To pull words from the same stem together, we applied snowball stemming, consistent with Fischer-Preßler et al. [2019]. To keep common compounds (e.g. "deutschsprachiger Einsteigerkurs" (Introductory course in German)) we keep all phrase of the length 2 occurring more than 1000 times in our corpora. Furthermore we cut all terms consisting of less than 2 letters and removing too special terms ($< 0.5\%$ of all documents) and too common terms (e.g. Volkshochschule, $> 30\%$ of all documents), which tell us less about the specific construction of a topic.

For the creation of our topic model we first try to select the optimal number of topics describing our corpus. For this step we compared different k within a range of $k_i = \{5, 6, 7, \dots, 30\}$ with regard to metrics like exclusivity, semantic coherence and residuals. We choose results $k_{res} = \{16, 17\} \in k_i$ over performing in all metrics with regard to the general mean for this metrics for a further evaluation. In the next step we applied a qualitative coding for all k_{res} to access which results yields the best humanly interpretative results.

For this step four coders labelled the topics according to Highest Probability, FREX, Score and Lift Words and checked the top 30 documents for each topic. Afterwards they indicated their safety of assignment on a scale from 1 (perfect fit) through 6 (maximal unsure). A list of the evaluated terms can be found in table 3 for the 16 topics model and in table 4 for the 17 topics model. To check differences in coding, we checked the summary statistics for the levels of uncertainty and obtained Mean = 2.11, SD = 0.89 and a Range from 1 to 5 for the model with 16 and Mean = 2.19, SD = 1.08 and a Range from 1 to 5 for the model with 17 topics. A visual comparison of the uncertainties can be obtained in Figure 6. Following this and a qualitative check of consistency between the topic labels, we selected the model with $k_{res} = 16$ for the further steps of our analysis. To utilise the topics in our embedding regressions (which only allow independent dummy variables at the moment, see also Rodriguez et al. [2021]) later on, topic proportions per course description were dichotomised by assigning the value 1 if a documents topic proportion for a chosen topic is over 0.25 and 0 if else. This cutoff point was selected because of the sharp right-skewed, leptocurtic distribution of the topic shares, indicating a good separation between courses dealing with a topic to some degree and courses without reasonable mentioning of a specific topic.

3.3 Word Embedding

Simplified, we can understand word embeddings as a dimension-reduced version of the co-occurrences of words with all other words in a specific corpus, represented in word-wise Vectors of D Dimensions (normally more than 100 to keep a relatively granular distinction). When looking at this vectors we can compute distance measures to compare their relative similarity in the whole embedding space, because we assume that words accompanying each other are more similar to one another than words who are rarely used in the same context. This is consistent with the distributional assumption of Firth [Rodriguez et al., 2021, Widdowson, 2007], stating that the general meaning of words can be accessed by their co-occurrence with other words. For our context, this seems somewhat appropriate. We would assume that courses teaching office skills include words like Word, Excel and table calculation and that these words should be relatively near in their meaning of a courses content. This strong focus on co-occurrences should also keep us somewhat safe when discovering synonyms (this should just be vectors with an extremely high level of similarity).

Due to this relatively simple and robust conceptual (not necessarily methodological) framework, word embeddings are useful in a variety of natural language processing tasks, e.g. linguistic shift in regard to gender and ethnicity stereotyping [Garg et al., 2018], changes in the media discourse [Azarbonyad et al., 2017], political texts [Gurciullo and Mikhaylov, 2017], and classes [Kozlowski et al., 2019].

While such models, (with many different implementations) offer many insights, inference about the structural factors behind a specific configuration of a word is difficult, because the position of every word is dependent on every other word. Theoretically, one could train different models for each realisation of a relevant (independent) variable (as in the case of Bianchi et al. [2020]), but this could become unfeasible for smaller subcorpora for each group [Rodriguez et al., 2021].

One possible approach to circumvent this problem are ALC embeddings. Here it is assumed that the content of a word w can be described by averaging the embedding of the word around it ($w(\mathbf{u}_w)$), multiplied by some relation-preserving (or weight) matrix A . This matrix can simply be calculated via regression which re-weights words according to the total sum of each word in the corpus [Rodriguez et al., 2021].

For our setup, we basically followed the proposed steps by Rodriguez et al. [2021] and additionally estimated a full GloVe embeddings model [Pennington et al., 2014] and a transformation matrix, because we assume specific words and combinations in the VHS dataset. One reason for this assumption is the specific Covid-19 situation at the time of data collection and the other is the specific focus of VHS offers to appeal to a large, often unspecific group of people. Following this assumption we estimated an embedding model with $D = 300$, a learning rate of 0.05 and a window size of 6. We then constructed our transformation matrix A via a feature (word) co-occurrence matrix with a weight of 1000.

3.4 Embedding Regression Models

À la carte Text Regression Models represent a relatively new branch of text-as-data applications combining word embeddings with the ability to include exogenous variables for a estimation of context specific parameter effects. The following notation is strictly orientated on the definitions by Rodriguez et al. [2021]. For our models we centred around a specific term (focal words, e.g. Bilanz) defined as $\mathbf{v}_w = A\mathbb{E}[\mathbf{u}_w]$ and this terms specific context defined by his near neighbours in the word embedding space. We can then calculate the resulting expectation, conditional on some covariate $X : \mathbb{E}[A\mathbf{u}_w | X]$ with our weighting Matrix A . Through this multidimensional representation of each included observation for the instance of an embedding \mathbf{v}_{w_i} our dependent variable \mathbf{Y} is not of length vector $n \times 1$, but rather $n \times D$ with $D = 300$ in our case [Rodriguez et al., 2021]. The usual matrix equation can then be formalised as:

$$\underbrace{\mathbf{Y}}_{n \times D} = \underbrace{\mathbf{X}}_{n \times p+1} \times \underbrace{\beta}_{p+1 \times D} + \underbrace{\mathbf{E}}_{n \times D} \quad (1)$$

with \mathbf{X} a matrix of p covariates and including a constant term and β as a set of p coefficients and an intercept (all of dimension D). Then \mathbf{E} is an error term. Similar to simple linear regression models, this allows us to estimate the effect of different variables on a dependent (in this case multivariate) variable \mathbf{Y} . For our case we used the previously specified vocabulary $V_{Business}$ as our set of dependent variables and the dichotomous variables p of "Course Pre-Requisites and Starting Information", "Basic digital skills", "Vocational and personal development", "Language course objectives in the CEFRL", "Refreshment and improvement of language skills", "Emotional Well-Being", "Basic and specialised commercial courses / accounting", "Sector-specific specialised courses", "IT / media basics / general apps", "Commercial IT / media applications", "Organisation and management", "Softskills", "Technical basic and specialised courses", "Technical IT / media applications" as independent variables in each regression model. The selected topics for p were chosen by their general fit with the business focus of our analysis (e.g. digital skills as necessary abilities for high-skill labour versus Dancing towards fitness as a more leisure oriented activity). Following this we specify our models i for the dictionary $V_{Business}$ as:

$$\underbrace{\mathbf{Y}_i}_{n \times D} = \underbrace{\mathbf{X}_i}_{n \times p+1} \times \underbrace{\beta_i}_{p+1 \times D} + \underbrace{\mathbf{E}_i}_{n \times D}; \forall \mathbf{Y}_i \in V_{Business} \quad (2)$$

In the following section we show some descriptive findings, discuss the results of the estimation process and evaluate on the predicted à la carte embedding results for the specific independent variables.

4 Results

In the first part of the results section we evaluate the distribution of the categories. The categories we describe here were self-chosen by the offering institution from a list that is divided into 6 categories which can be differentiated into roughly 80 subcategories. A given subcategory (e. g. "English") hereby belongs exclusively to one specific category (e.g. "Languages").

Regarding the categories, "Health" is assigned to roughly one third (28.07 %) of the offers in the course-finder. It is followed by "Languages" (19.95 %), "Culture - Creative" (18.89 %), and "Work - Job" (17.43 %), which were categorised relatively evenly. They are followed by "Politics - Society - Environment" (14.38 %), the category one in every seven offers is in. Rarely, meaning less than 1 per cent, the categories of "School - Graduation" (0.69 %) and "Basic Education" (0.59 %) were assigned.

This relatively rough categorisation already allows us to draw some preliminary conclusions: The course offers of the VHS are mostly situated in activities of self-improvement (health, languages). Nevertheless, further education doesn't loose ground here: there are a lot of courses in the part of "training along the job". On the other hand, courses that replace or make up for the lack of basic schooling ("School-Graduation" and "Basic Education") are relatively rare, maybe because those offers are often accompanied by a social (re-)integration, placing them more into the responsibility of social welfare.

In the following, the subcategories will be focused. The three subcategories, which account for most of the offers are "Exercise / Fitness" (9.26 %), "IT / media basics / general apps" (9.12 %), as well as "Relaxation / Stress management" (8.66 %). Especially the first and the latter focus on leisure activities, whereas the second helps keeping up with the latest technical developments or to acquire these fundamentally. An other huge part of the offers of VHS are language courses the one among them that is categorised the most, is "German as a foreign / second language" (7.47 %), followed by "English" (3.66 %), "Spanish" (2.28 %), "Italian" (1.51%), and "French" (1.39%). Apart from that, 20 further foreign languages, dialects, and German Sign Language (and a rest-category) are offered. Here one can see the macro-didactic consideration of the necessity for courses, that are financially beneficial. Apparently the demand for sports / health courses and language courses is relatively high and stable. A special case here are the courses "German as a foreign / second language", as they are compulsory for some migrants (e.g. asylum seekers, holder of some visa, EU-Migrants) and often financed and / or subsidised by federal or regional agencies [BAMF, 2020].

While "IT / media basics / general apps" is already one of the sub-categories, that will be focused in the following, and relatively prevalent, "Basic and specialised commercial courses / accounting" (1.61 %) and "Soft skills / job application training" (1.26 %) are overall relatively rare, but in the field, we focus on they are considerably important. The further sub-categories in our field of interest are "Organisation / Management" (0.99 %), "Technical IT / media applications" (0.89 %), "Commercial IT / media applications" (0.76 %), "Sector-specific specialised courses" (0.69 %), and "Technical basic and specialised courses" (0.25 %). Although the latter might seem irrelevant, due to the high number of courses the VHS offer, there are still 150 courses within the least chosen category, that seems to match our field of interest. In general, the ratios of categories, as well as subcategories, of the courses in our dataset in general match with the results presented by the VHS-statistics concerning the previous year of 2019 Huntemann et al. [2019]. Another relevant separation of the courses is via its topics. As the construction of the topic model has been explained above, in the following its results will be described (see Table 1). It is eye-catching that the most assigned category is "Physical health and mindfulness" (17.27 %). It is followed by "Handicraft Culinary Courses" (15.59 %). 11 % of texts contain more than 25 % of its content from the topics of "Basic digital skills" and "Emotional Well-Being", respectively. The last topic that accounts for more than 25 % of the words in one in 10 texts is "World Knowledge".

After the mere presentation of the distribution of the (sub-)categories and the topics, in the following (Figure 6) they were combined. On the bottom the categories are annotated. As one course can just be assigned to one subcategory they exclude each other. On the left all 16 topics are aligned. Here, one course-description can be assigned to several topics. The grid shows the percentages within the column e.g. which share of course-descriptions was assigned a given topic under the condition of a pre-chosen category. This means, that within the category of Work - Job 55 % of course descriptions have been assigned to the topic of "Basic digital skills", meaning VHS offer a variety of classes focusing the societal challenge of digitalisation within that category. Regarding the Health-category the topic "Physical health and mindfulness" is very dominant herein (58 %). The topics of "Handicraft Culinary Courses", which include healthy and dietitian cooking as well, just like "Emotional Well-Being", which is a part of an all-embracing health, seem to be

Table 1: Distribution of the Topics in the final 16 topics model

Topic ($N = 53942$)	Mean	SD	Skewness	Kurtosis
Physical health and mindfulness	0.1727	0.3780	1.7316	0.9986
Course Pre-Requisites and Starting Information	0.0522	0.2225	4.0247	14.1985
Dancing towards fitness	0.0344	0.1822	5.1119	24.1329
Basic digital skills	0.1110	0.3141	2.4772	4.1367
Citizenship course	0.0263	0.1599	5.9242	33.0979
Remote Language Courses	0.0057	0.0756	13.0775	169.0280
General German Integration Course	0.0190	0.1366	7.0425	47.5990
Integration Course	0.0037	0.0608	16.3318	264.7384
Vocational and personal development	0.0180	0.1329	7.2548	50.6334
Basic language Courses	0.0721	0.2586	3.3089	8.9490
Handicraft & Culinary Courses	0.1559	0.3628	1.8971	1.5990
Language course objectives in the CEFRL	0.0163	0.1268	7.6320	56.2501
Refreshment and improvement of language skills	0.0550	0.2279	3.9054	13.2524
Emotional Well-Being	0.1166	0.3209	2.3900	3.7123
World Knowledge	0.0989	0.2985	2.6872	5.2213
Advanced Language Courses	0.0048	0.0691	14.3278	203.2940

very important, too (17 and 11 % respectively). For the category "Culture - Creative" the topic of "Handicraft Culinary Courses" is a great share (49 %). This is not surprising, as handicrafts often are creative (e.g. sewing), and many culinary courses focus on local cuisine (e.g. Pizza and Tapas). For the category "Politics - Society - Environment", 42 % of courses are assigned with the topic of "World Knowledge", while "Emotional Well-Being" (29 %) is the other main topic here. Given those two dominant topics, all other 14 topics seem to be relatively irrelevant within this category (less than 10 per cent each). Regarding the category of "Languages", the most important topics here are "Basic language Courses" (36 %) and "Refreshment and improvement of language skills" (18 %). This accounts for the fact, that a lot of these courses focus on aspects such as communication and social inclusion, rather than achieving a high competence in a foreign language quickly. The categories of "Basic Education" and "School - Graduation" who, as pointed out above, include the least courses, seem not to have a predominant topic (the closest would be the "Refreshment and improvement of language skills" for "School - Graduation", which is not surprising as German and a foreign language are relevant for either graduation in Germany [Kultusministerkonferenz, 2020], none other topic above 15 %). Unsurprisingly the share of topics by categories showed us, that certain of each seem to be linked together by the content of the course descriptions (e. g. "Work - Job" and "Basic digital skills"). Further it allows us to state, that we chose our topics wisely, as none of the findings here surprises us dramatically. Ultimately some combinations seem to be rather exclusive (e.g. "Work-Job" and "Basic digital skills"), while other topics seem to be relevant to a certain point in every category (e. g. "Emotional Well-Being" and "Course Pre-Requisites and Starting Information").

In a next step, we evaluate on the results of the multivariate, multiple regressions models for the terms $V_{Business}$, presented in Table 2. Following [Rodriguez et al., 2021] we discuss the relative difference of the estimates and not the absolute values of $\hat{\beta}$. In general, estimates for $\hat{\beta}$ show very little differences between the used terms. Relatively low contextual separation can be obtained for the subcategories "Organisation and Management" ($\hat{\beta}$ between 0.041 – 0.044), "Sector-specific specialised courses" ($\hat{\beta}$ between 0.042 – 0.045), "softskills" ($\hat{\beta}$ between 0.042 – 0.044) and for the topic "Course Pre-Requisites and Starting Information" ($\hat{\beta}$ between 0.054 – 0.056). Subcategory-wise, this could indicate that such categorisations tend to consist of a more plural composition in regard to the evaluated vocabulary than other categories. One possible explanation for the low value of "Course Pre-Requisites and Starting Information" could also lead in the same direction, stating that general course information are a general part of course descriptions. While the last point is rather trivial, we also observe higher levels of separation for subcategories like "Technical IT / media applications" ($\hat{\beta}$ between 0.098 – 0.114), "Basic and specialised commercial courses / accounting" ($\hat{\beta}$ between 0.069 – 0.072) and the topic "Language course objectives in the CEFRL" ($\hat{\beta}$ between 0.101 – 0.104) indicating a more specific usage of the terms of our $V_{Business}$. All coefficients are significant on the $p < 0.05$, showing statistically significant separation of the chosen categories in our regression models.

While the above mentioned differentiation of terms according to subcategories and topics can give us some indication of factors of separation, it does not show the differences in the textual representation on a word level. A display of the nearest neighbours on such is shown in figure 3. Here we present the results of a bootstrapped, permuted selection of similar terms via cosine similarity. For our showcase we selected six combinations of the potentially 14×25 combinations between p and elements of $V_{Business}$. First, we look at "Basic and specialised commercial courses /

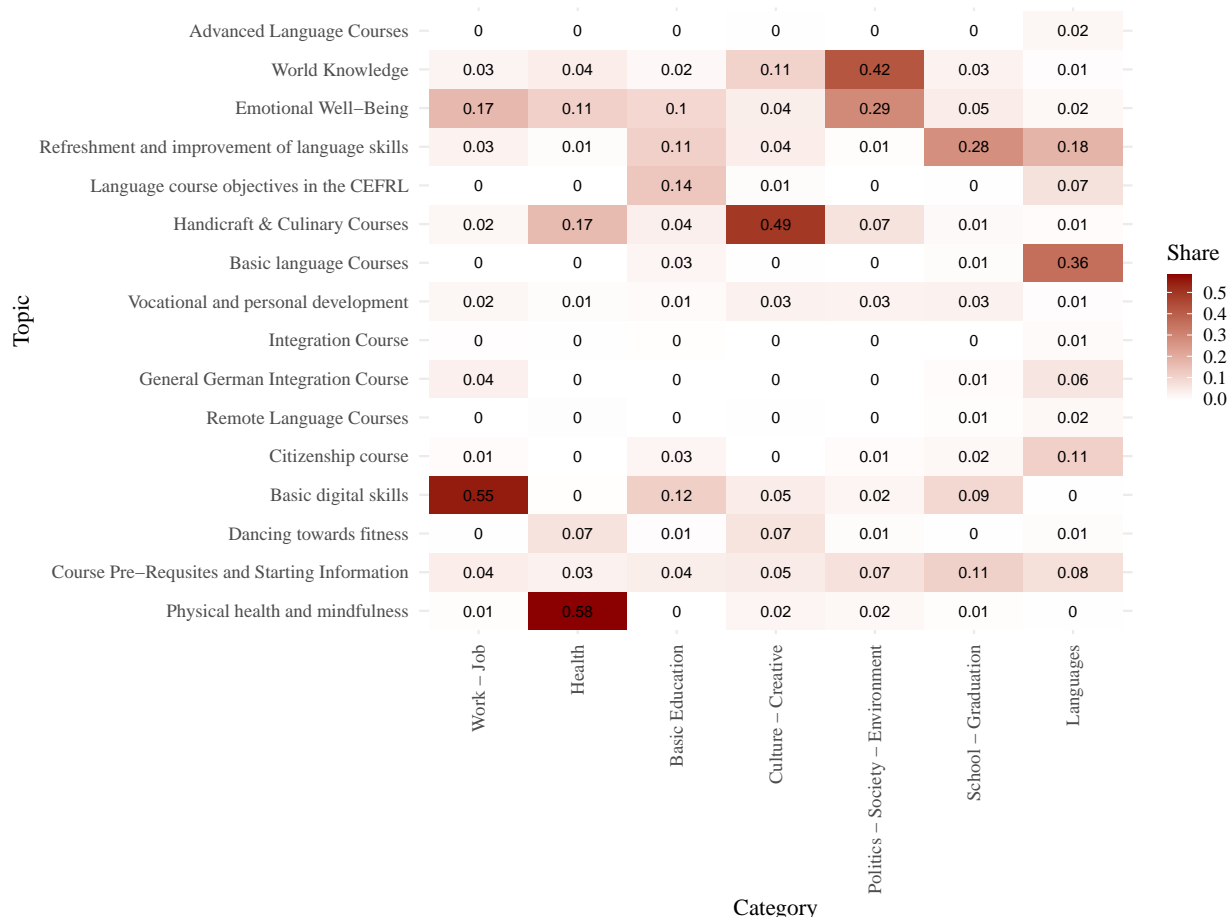


Figure 2: Share of Topic by Category

accounting Bilanz" (balance sheet). It is observable that terms like "Inventur" (stocktaking), "Inventar" (inventory), words with the stem "bilanz" (balance), "Konten" (accounts) and "Rückstellung" (provision) show the contextual usage of "Bilanz" (balance sheet) as a term used in specialised courses dealing with banking and account. In this framing the selected term is dominantly (and statistical significant on the $p < 0.05$ -level) connected to such forms of specialisation and less used in more basic forms. Comparing these results with the nearest neighbours, if the course description consists of minimal 25 % "Pre-Requisites and Starting information", specific words like "Rückstellung" (provision) stay nearby, but terms like "Grundlagen" (fundamentals) and "Inhalte" (contents) show a bigger orientation on basic information.

Regarding the term of "Buchführung" (accounting), the subcategories "Basic and specialised commercial courses / accounting" and "Sector-specific specialised courses" have been analysed. The former points out the relevance of the terms of "Inventur" (stocktaking) and "Buchführung" (accounting) itself. This might, on one hand, point out, that this term is very specific, and, on the other hand, might indicate, that potential learners in this field look for the explicit term. Other relevant words in this context are "Bilanz" (balance sheet) and "ordnungsgemäßer" (proper), which implicits accuracy. Overall, all words in this combination are distinct. Focusing the next terms "Grundsätze" (principles), "einstiegende" (introductory) and "Grundlagen" (fundamentals), they reveal a basic level of the courses, according to their description. This seems to be especially interesting, if this is compared to the word of "Bilanz" (balance sheet), ranging in the same subcategory, where none of this words appears. This might lead to the conclusion, that "Buchführung" (accounting) is a more basic task than "Bilanz" (balance sheet), or at least it can be taught with less previous knowledge.

The next combination, that will be focused is that of sector-specific specialised courses - Buchführung (accounting). The first thing that is eye-catching are the comparably big confidence-intervals. The first to words here being "Grundkenntnisse" (basic knowledge) and "Voraussetzungen" (prerequisites) lead to the conclusion that a basic accounting knowledge is continued and broadened in these courses. Specific contents hereby seem to be computer

Table 2: Summary of of the parameter estimates $\hat{\beta}$ for the whole vocabulary

Coefficients ($N = 53942$)	Type	Min Estimate	Max Estimate	Min Std.	Max Std.	P - Value
Course Pre-Requisites and Starting Information	Topic	0.0541	0.0560	0.0004	0.0007	<0.05
Basic digital skills	Topic	0.0622	0.0654	0.0004	0.0009	<0.05
Vocational and personal development	Topic	0.0522	0.0548	0.0004	0.0008	<0.05
Language course objectives in the CEFRL	Topic	0.1008	0.1042	0.0005	0.0012	<0.05
Refreshment and improvement of language skills	Topic	0.0588	0.0605	0.0003	0.0006	<0.05
Emotional Well-Being	Topic	0.0459	0.0470	0.0002	0.0004	<0.05
Basic and specialised commercial courses / accounting	Subcategory	0.0685	0.0722	0.0006	0.0012	<0.05
Sector-specific specialised courses	Subcategory	0.0418	0.0447	0.0006	0.0011	<0.05
IT / media basics / general apps	Subcategory	0.0442	0.0473	0.0006	0.0010	<0.05
Commercial IT / media applications	Subcategory	0.0539	0.0609	0.0007	0.0014	<0.05
Organisation and management	Subcategory	0.0411	0.0435	0.0006	0.0010	<0.05
Softskills	Subcategory	0.0415	0.0440	0.0005	0.0008	<0.05
Technical basic and specialised courses	Subcategory	0.0543	0.0587	0.0008	0.0018	<0.05
Technical IT / media applications	Subcategory	0.0984	0.1142	0.0033	0.0055	<0.05

assistance ("EDV" (EDP), "Datenverarbeitung" (data processing) and "Windows") as well as "Geschäfte" (businesses). In comparison to the former described combination, the latter seems to be its continuation.

Regarding "Basic and specialised commercial courses / accounting - Management" the most important terms are "Kosten" (costs) and "Wirtschaftlichkeit" (economic efficiency). Which makes economic survival a relevant part of the understanding of management by the VHS. Investment ("Investition" and "Investitionen") is a part of that, too, and a relevant term here as well. Unsurprisingly the term of "Unternehmensführung", the top level of management in German, appears here as well, which indicates that both terms seem to be used synonymous in the course description. In comparison to the other terms within this subcategory we can observe, that none of the words from there is close to Management on the 10 % level, which leads us to the conclusion, that neither balancing nor accounting seem to be tasks of the management-level, at least when it comes down to learning.

The last combination that will be presented here is "Course Pre-Requisites and Starting Information - Daten (Data)". As important terms here, "intern" (internal), "unterliegen" (underlie) and "Datenschutz" (data protection), are relevant information, transmitted via the course description here is the confidentiality of the used data. The personal data protection of participants seems to be relevant as well (Terms of "mobilen" (mobile) and "Cloud"), which might lead to the conclusion, that this refers to hand-on-courses about individual data privacy, especially on mobile devices. Further, in comparison to "Bilanz" (balance sheet) within the same topic, "Daten" (data) seems to focus more on the willingness to conspire on the data used, than on actual (preliminary) competencies.

In general the near neighbours of the topics, respectively subcategory and the terms illustrated us, that all of the focused words seem to be embedded in specifically constructed sentences, which differentiate their grouping either by the VHS themselves or us as the analysing scientists.

5 Discussion

Following the results of the analysis of sub(categories), topics and context of selected terms, we observe a high level of heterogeneity in the structure of course offers. While we are able to reproduce the majority of the findings in regard to relative differentiation of course in categories and subcategories [Huntemann et al., 2019], we detect fundamental textual schemes via topic modelling which in some cases align with the mapping by categories and in some cases diverge.

Here, we observe cross-category schemes like general course information, basic digital skills, general knowledge and emotional well-being as patterns in the textual representation of course offers. One result of this is that (most) textual

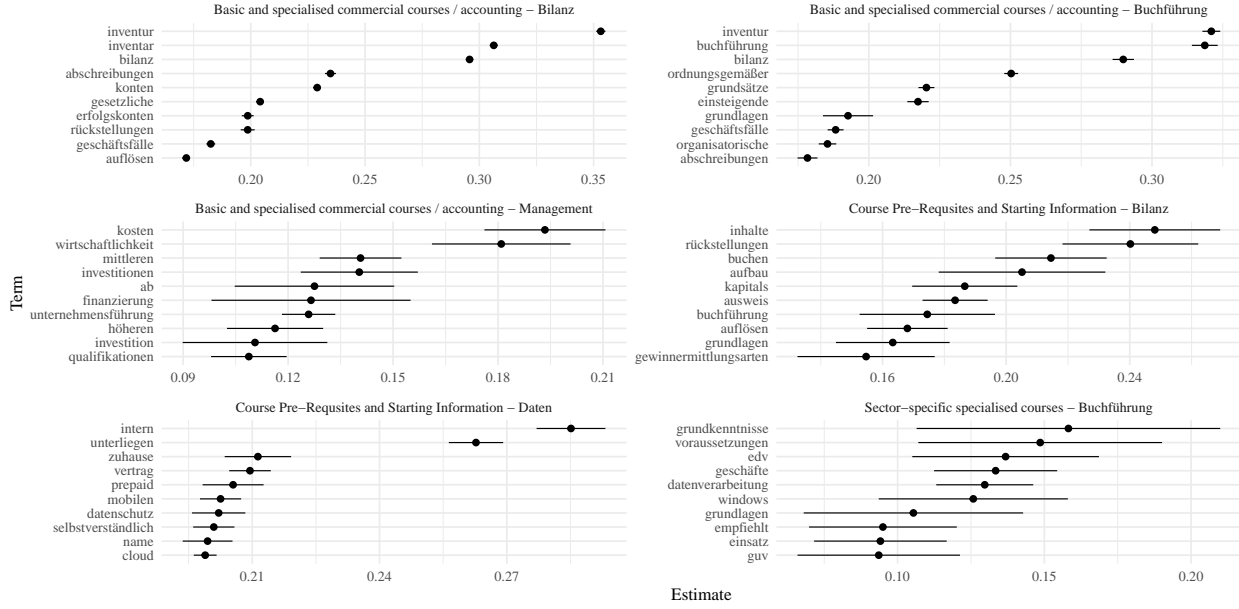


Figure 3: Near neighbours for the combination of Topic/Subcategory and Term, all effects are significant on the $p = 0.05$ level.

representations of course descriptions are not solely working as an assemblage of curricular content and keywords, but rather as forms of structural framing and narrative creation with co-occurrences like work, digital basics, and emotional well being pointing towards a socially and personal embedded view of work-related courses. This is consistent with previous discussion by Groth [2018].

Lines of differentiation are also observable looking at a more granular level on textual contexts via word embedding regression modelling and near neighbours via ALC embeddings. Content-related separations of work-specific terms are ubiquitous, indicated by the significant differences between categories and topics via topic regression models. While these models show a general tendency, we were able to illustrate on an exemplary level, that topic-wise classified basic-level-courses may offer elementary knowledge about the contexts, while sector-specific courses set the focus on more technical aspects. Another point is that structurally related terms are different in the embeddedness in the prior knowledge. Here we observe, on a exemplary, level that naming matters and related terms can signal different content levels to interested persons. As illustrated, VHS have the economic necessity to drag participants into their courses. By addressing them adequately, they can facilitate themselves this task. Therefore, we resume specific course-description, which are interpreted as addressing. An illustrative example here would be the connection to the Xpert Business-System, the VHS of Baden-Württemberg developed and that is internally used to classify and separate different topics and levels of economic and administrative knowledge [Xpert Business, 2021]. A linkage to that systematisation in the course description might motivate employers for their employees or themselves to favour for a certain course, as the targeted outcome is measurable.

6 Conclusion

By leveraging innovative text-as-data methods we were able to take a deeper dive in the structural compositions of online publicly available VHS course offers in Germany at the turn of the year 2020/2021. Further research is needed to evaluate on a broader level about the course contents and general embeddedness in the German educational system. While we were able to evaluate on a textual level about topics and selected terms, we did not gather information about potentially relevant background factors like location, specific needs, pricing, equipment and limitation factors (personal and structural). Further steps could be an integration of such information in a strict, causal orientated framework to evaluate the multi-causal nature of course offer creation.

Another potentially relevant addition could be integration of multi-language aware natural language processing applications to co-evaluate non-german text with a complete model. While this is especially interesting for language courses, other categories like english-only business courses could also be of interest for the analysis of language- and level-wise differences. This could include in particular sector-specific technical language courses in the analysis, which

are located in the overlap area of the categories "Language" and "Work - Job" [Deutscher Volkshochschul-Verband e.V., 2021]. Another critical aspect is the definition of context solely by co-occurrences. While this enables relatively easy computation and analysis, context means more. On the one hand, other approaches like narrative mining [Ash et al., 2021] or contextualisation of different entities of meaning [Brand et al., 2021] can help to narrow the gap of understanding from an computational perspective. On the other hand, more qualitative approaches could be a useful addendum to the research process, emphasising the complex and subjective creation process of course planing and its representation in the VHS-Kursfinder, which has not been done so far [Giseke et al., 2018].

References

- Uwe Gartenschlaeger. Introduction: Community Learning centres as spaces for Lifelong Learning in the 21st Century. In *Adult education centres as a key to development - challenges and success factors*, pages 8–16. DVV International, Bonn, 2017.
- Shalini Singh and Soren Ehlers. Employability as a global norm: comparing transnational employability policies of OECD, ILO, World Bank Group, and UNESCO. In *Studies on Adult Learning and Education*, pages 131–148. University Press, Firenze, 2020.
- Heribert Hinzen. 100 years of Volkshochschule -50 years of DVV International. Local and global perspectives on adult education and lifelong learning. In *Studies on Adult Learning and Education*, pages 199–212. University Press, Firenze, 2020.
- Rolf Arnold, Henning Pätzold, and Mario Ganz. Weiterbildung und Beruf. In *Handbuch Erwachsenenbildung / Weiterbildung*, pages 931–946. Springer VS, Wiesbaden, 6. auflage edition, 2018.
- Kommission der Europäischen Gemeinschaften. Memorandum über Lebenslanges Lernen. Arbeitsdokument der Kommissionsdienststellen 1832, EU-Kommission, Brüssel, 2000. URL https://www.hrk.de/uploads/tx_szconvention/memod.pdf.
- Vanna Boffo. Storytelling and other skills: Building employability in higher education. In *Studies on Adult Learning and Education*, pages 31–50. University Press, Firenze, 2020.
- Maja Avramovska, Esther Hirsch, and Beate Schmidt-Belau. *Adult education centres as a key to development - challenges and success factors*. DVV International, Bonn, 2017. URL https://www.dvv-international.ba/fileadmin/files/bosnia-and-herzegovina/Documents/IPE-78_web.pdf#page=47.
- Hans-Georg Kotthoff. Between excellence and equity: the case of the german educational system. *Revista Española de Educación Comparada*, 18(2011):27–60, 2011.
- Rita Süßmuth and Karl Heinz Eisfeld. Volkshochschule. In *Handbuch Erwachsenenbildung / Weiterbildung*, pages 763–784. Springer VS, Wiesbaden, 6. auflage edition, 2018.
- Xpert Business. Unser Leitbild, 2021. URL <https://www.xpert-business.eu/de/leitbild/index.html>.
- Klaus Heuer and Nicole Kuprian. Anchoring adult education centres in society. In *Adult education centres as a key to development - challenges and success factors*, pages 153–176. DVV International, Bonn, 2017.
- Nella Huntemann, Nicolas Echerti, Thomas Lux, and Elisabeth Reichart. Volkshochschul-Statistik. DIE Survey 58, DIE, Bonn, 2019. URL <https://www.die-bonn.de/weiterbildung/statistik/vhs-statistik>.
- Jutta Reich-Claassen and Aiga von Hippel. Programm- und Angebotsplanung in der Erwachsenenbildung. In *Handbuch Erwachsenenbildung / Weiterbildung*, pages 1403–1424. Springer VS, Wiesbaden, 6. auflage edition, 2018.
- Wiltrud Giseke, Aiga von Hippel, Maria Stimm, Iva Georgieva, and Stephanie Freide. Programmarchive und -sammlungen der Erwachsenenbildung / Weiterbildung. In *Handbuch Erwachsenenbildung / Weiterbildung*, pages 451–474. Springer VS, Wiesbaden, 6. auflage edition, 2018.
- Annette Groth. Die Rolle der Beruflichen Bildung an Volkshochschulen. *Hessische Blätter für Volksbildung*, 04: 358–364, 2018. URL <https://elibrary.utb.de/doi/10.3278/HBV1804W371>. Publisher: wbv Publikation.
- Jörg Dittrich. Berufliche Bildung als Gestaltungshilfe aktiver Berufsbiografie. *Hessische Blätter für Volksbildung*, 04: 368–370, 2018. URL <https://elibrary.utb.de/doi/10.3278/HBV1804W368>. Publisher: wbv Publikation.
- Peter Faulstich. Weiterbildung und Technik. In *Handbuch Erwachsenenbildung / Weiterbildung*, pages 947–972. Springer VS, Wiesbaden, 6. auflage edition, 2018.
- Deutscher Volkshochschul-Verband e.V. Vielfalt und Kompetenz für Erwachsenenbildung, 2021. URL https://www.volkshochschule.de/verbandswelt/programmbereiche/sprache_und_integration/index.php.
- ZBW. Standard-Thesaurus Wirtschaft, October 2021. URL <http://zbw.eu/stw/version/latest/about>.

- Diana Fischer-Preßler, Carsten Schwemmer, and Kai Fischbach. Collective sense-making in times of crisis: Connecting terror management theory with twitter user reactions to the berlin terrorist attack. *Computers in Human Behavior*, 100:138–151, 2019.
- David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022, 2003.
- Christopher A Bail, Taylor W Brown, and Marcus Mann. Channeling hearts and minds: Advocacy organizations, cognitive-emotional currents, and public conversation. *American Sociological Review*, 82(6):1188–1213, 2017.
- Laura K Nelson. Computational grounded theory: A methodological framework. *Sociological Methods & Research*, 49(1):3–42, 2020.
- Carsten Schwemmer. The limited influence of right-wing movements on social media user engagement. *Social Media+ Society*, 7(3):20563051211041650, 2021.
- Pedro L Rodriguez, Arthur Spirling, and Brandon M Stewart. Embedding regression: Models for context-specific description and inference. 2021.
- Henry Widdowson. Jr firth, 1957, papers in linguistics 1934–51. *International Journal of Applied Linguistics*, 17(3): 402–413, 2007.
- Nikhil Garg, Londa Schiebinger, Dan Jurafsky, and James Zou. Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, 115(16):E3635–E3644, 2018.
- Hosein Azarbondy, Mostafa Dehghani, Kaspar Beelen, Alexandra Arkut, Maarten Marx, and Jaap Kamps. Words are malleable: Computing semantic shifts in political and media discourse. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1509–1518, 2017.
- Stefano Gurciullo and Slava J Mikhaylov. Detecting policy preferences and dynamics in the un general debate with neural word embeddings. In *2017 International Conference on the Frontiers and Advances in Data Science (FADS)*, pages 74–79. IEEE, 2017.
- Austin C Kozlowski, Matt Taddy, and James A Evans. The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84(5):905–949, 2019.
- Federico Bianchi, Valerio Di Carlo, Paolo Nicoli, and Matteo Palmonari. Compass-aligned distributional embeddings for studying semantic differences across corpora. *arXiv preprint arXiv:2004.06519*, 2020.
- Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.
- BAMF. Trägerrundschreiben Integrationskurse, December 2020. URL https://www.bamf.de/SharedDocs/Anlagen/DE/Integration/Integrationskurse/Kurstraeger/Traegerrundschreiben/2020/traegerrundschreiben-23_20201210.pdf?__blob=publicationFile&v=3.
- Kultusministerkonferenz. Bildungsstandards der Kultusministerkonferenz, 2020. URL <https://www.kmk.org/themen/qualitaetssicherung-in-schulen/bildungsstandards.html>.
- Elliott Ash, Germain Gauthier, and Philine Widmer. Text semantics capture political and economic narratives. *arXiv preprint arXiv:2108.01720*, 2021.
- Alexander Brand, Tim König, and Wolf J. Schünemann. Using heterogenous information networks for integrative discourse mapping: The covid19 example. In *1st International Workshop on Knowledge Graphs for Online Discourse Analysis, KnOD 2021*, 2021.

7 Appendix

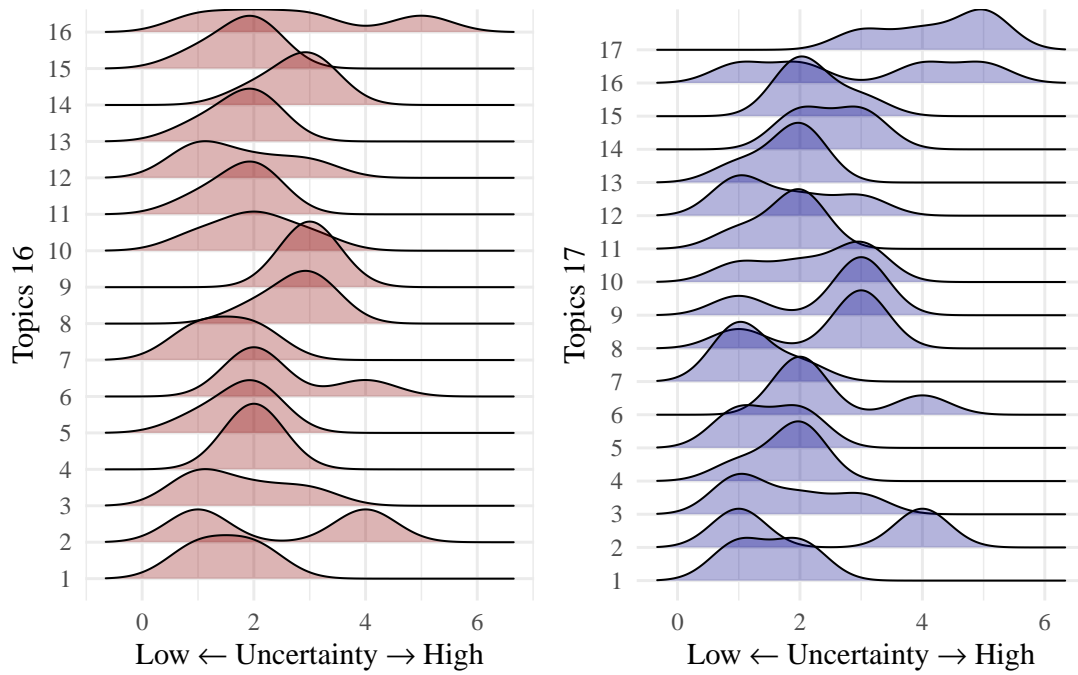


Figure 4: Ridgeline plots of the uncertainty ratings between the 16 and 17 Topic Models

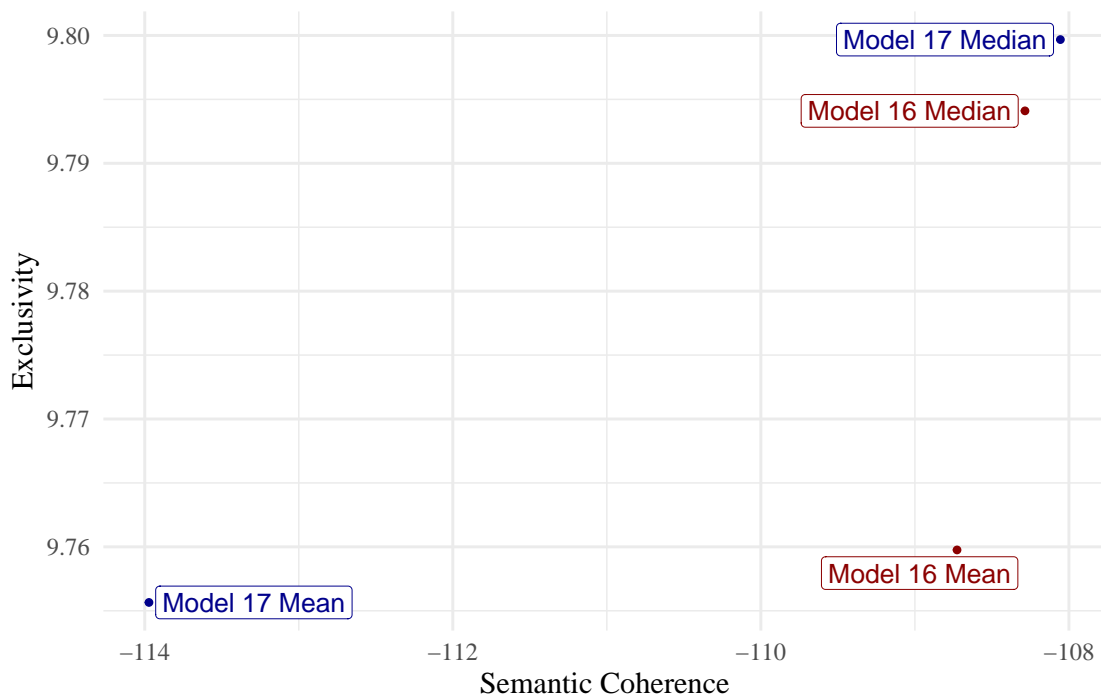


Figure 5: Coherence/Exclusivity plots between the 16 and 17 Topic Models

Table 3: Terms for the labelling of the Topic Models, 16 Topics Model

Topic	Type	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7
1	prob	körper	übungen	yoga	entspannung	beweglichkeit	geist	training
1	frex	yoga	decke	rücken	pilates	wirbelsäule	verspannungen	beweglichkeit
1	lift	bauch-	dehn-	yogamatte	yoga	asanas	atemübungen	bewegungsübungen
1	score	gesunden	körper	yoga	beweglichkeit	entspannung	pilates	geist
2	prob	bitte	anmeldung	vhs	uhr	erhalten	ab	30
2	frex	30	00	beratung	uhr	per	eur	zoom
2	lift	beratung	00	headset	internetverbindung	webcam	30	mikrofon
2	score	beratung	uhr	anmeldung	00	vhs	30	eur
3	prob	spaß	musik	macht	einfach	bewegung	freude	ganz
3	frex	spaß	musik	zumba	tanzen	fit	spielen	frauen
3	lift	plus	rhythmen	zumba	tanzen	tanz	spaß	rhythmus
3	score	plus	spaß	musik	zumba	tanzen	bewegung	tanz
4	prob	erstellen	umgang	lernen	arbeiten	funktionen	inhalte	internet
4	frex	excel	word	tabellen	apps	funktionen	windows	smartphone
4	lift	siehe	apps	excel	formatieren	formeln	smartphones	speichern
4	score	siehe	excel	erstellen	tabellen	word	windows	smartphone
5	prob	prüfung	test	fragen	deutschland	anmeldung	modul	flüchtlinge
5	frex	test	flüchtlinge	einbürgerungstest	modul	prüfung	anmeldeschluss	bundesamt
5	lift	33	einbürgerungstest	flüchtlinge	test	bamf	dauert	prüfungen
5	score	flüchtlinge	prüfung	test	einbürgerungstest	bundesamt	bamf	zuwanderer
6	prob	sollten	dass	dafür	drei	termine	kurse	online
6	frex	sollten	dafür	termine	notwendig	gern	drei	aufgrund
6	lift	buchbar	aufeinander	aufbauen	einzeln	passenden	sollten	aufgrund
6	score	buchbar	termine	sollten	dafür	einzeln	online	angebot
7	prob	unterrichtsstunden	sowie	100	deutschkenntnis	deutschland	europäischen	geschichte
7	frex	unterrichtsstunden	xpert	deutschkenntnis	zertifikat	business	europäischen	100
7	lift	abschlusstests	lid	43	ämter	alltagsbezogenen	skalierten	themenbereiche
7	score	lid	unterrichtsstunden	integrationskurs	xpert	deutschkenntnis	orientierungskurs	rechtsordnung
8	prob	möglich	müssen	durchgeführt	ermäßigung	abschluss	migration	personen
8	frex	ermäßigung	migration	durchgeführt	möglich	abschluss	zusammenarbeit	grundlage
8	lift	migration	ermäßigung	erster	zusammenarbeit	leider	grundlage	selbstständig
8	score	migration	möglich	ermäßigung	zusammenarbeit	müssen	durchgeführt	kursgebühr
9	prob	sowie	ersten	erhalten	grundlagen	beim	teil	themen
9	frex	ersten	vorgestellt	praxis	liegt	bekannt	thema	besprochen
9	lift	gegeben	bekannt	theorie	kurstag	ersten	theoretischen	vorgestellt
9	score	gegeben	ersten	grundlagen	praxis	thema	vorgestellt	einführung
10	prob	sprechen	lehrbuch	isbn	a1	grammatik	sprache	lernen
10	frex	lektion	grammatik	klett	sprechen	sprachen	lehrwerk	lehrbuch
10	lift	sprachen	href	lehrwerk	spanisch	klett	restaurant	einkaufen
10	score	sprachen	isbn	lektion	a1	verlag	lehrbuch	klett
11	prob	eigene	lernen	verschiedene	bitte_mitbringen	mitbringen	gemeinsam	eigenen
11	frex	materialkosten	küche	entstehen	kursleiterin	nähmaschine	zutaten	material
11	lift	gemüse	ii	leinwand	schürze	verbrauch	zutaten	abgerechnet
11	score	ii	nähmaschine	küche	materialkosten	zutaten	nähen	schürze
12	prob	einfache	schreiben	ende	verstehen	lesen	stufe	texte
12	frex	schreiben	verstehen	kurze	einfache	ende	lesen	sätze
12	lift	entspricht	vertraute	sätze	wörter	kurze	wesentliche	schreiben
12	score	entspricht	einfache	schreiben	stufe	verstehen	vertraute	sätze
13	prob	innen	möchten	teilnehmer	vorkenntnisse	bereits	vermittelt	kenntnis
13	frex	möchten	innen	bereits	erweitern	atmosphäre	teilnehmende	kurs_richtet
13	lift	geringen	auffrischen	gedacht	gelernt	englischen	kleingruppe	kurzer
13	score	geringen	innen	teilnehmer	möchten	vorkenntnissen	erweitern	kurs_richtet
14	prob	lernen	mehr	menschen	kinder	erfahren	dabei	eigenen
14	frex	erkennen	kinder	eltern	seminar	besser	kindern	stimme
14	lift	referenzrahmen	auftreten	erkennen	konflikte	stimme	angst	strategien
14	score	referenzrahmen	seminar	kinder	stimme	erkennen	eltern	methoden
15	prob	vortrag	welt	jahren	gibt	seit	immer	zeit
15	frex	vortrag	heute	jahren	wurde	welt	warum	livestream
15	lift	südost	professor	ausgezeichnet	jahrhundert	europa	zuletzt	vortrag
15	score	südost	vortrag	heute	professor	livestream	jahren	warum
16	prob	teilnehmenden	möglichkeit	erforderlich	statt	hilfe	stellen	vhs.cloud
16	frex	teilnehmenden	vhs.cloud	kurs_findet	kursleitung	möglichkeit	digitalen	absprache
16	lift	unterrichtsmaterialien	weitergeführt	kursleitenden	präsenzunterricht	ergänzung	kursleitung	kurs_findet
16	score	unterrichtsmaterialien	digitalen	vhs.cloud	teilnehmenden	kurs_findet	statt	möglichkeit

Table 4: Terms for the labelling of the Topic Models, 17 Topics Model

Topic	Type	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7
1	prob	körper	übungen	yoga	entspannung	beweglichkeit	geist	alltag
1	frex	pilates	yoga	decke	rücken	verspannungen	wirbelsäule	beweglichkeit
1	lift	dehn-	entspannungsfähigkeit	gesunden	hatha-yoga	körperhaltungen	pilates	wohltuenden
1	score	gesunden	körper	yoga	beweglichkeit	entspannung	pilates	geist
2	prob	anmeldung	bitte	uhr	vhs	statt	erhalten	30
2	frex	30	uhr	00	beratung	eur	per	16
2	lift	beratung	00	30	headset	13	internetverbindung	eur
2	score	beratung	uhr	anmeldung	00	30	eur	vhs
3	prob	spaß	musik	macht	bewegung	einfach	freude	frauen
3	frex	spaß	musik	tanzen	schritte	zumba	fit	tanz
3	lift	plus	zumba	tanzen	rhythmen	tanz	rhythmus	spaß
3	score	plus	spaß	musik	zumba	tanzen	bewegung	tanz
4	prob	erstellen	umgang	lernen	arbeiten	funktionen	internet	inhalte
4	frex	excel	tabellen	word	funktionen	apps	windows	smartphone
4	lift	siehe	formatieren	tabellen	dateien	einfügen	grafiken	formeln
4	score	siehe	excel	erstellen	tabellen	word	windows	smartphone
5	prob	test	fragen	deutschland	modul	flüchtlinge	anmeldung	prüfung
5	frex	test	flüchtlinge	einbürgerungstest	modul	deutsche	33	bamf
5	lift	flüchtlinge	33	einbürgerungstest	test	bamf	dauert	prüfungen
5	score	flüchtlinge	test	einbürgerungstest	bundesamt	bamf	orientierungskurs	deutschland
6	prob	sollten	dafür	drei	dass	termine	online	kurse
6	frex	sollten	dafür	termine	gern	drei	aufgrund	notwendig
6	lift	buchbar	aufeinander	aufbauen	einzel	passenden	aufgrund	sollten
6	score	buchbar	termine	sollten	einzel	dafür	online	module
7	prob	prüfung	unterrichtsstunden	sowie	deutschkenntnisse	100	deutschland	geschichte
7	frex	xpert	unterrichtsstunden	deutschkenntnisse	business	business	allgemeine	100
7	lift	lid	43	ämter	alltagsbezogenen	abschlusstests	xpert	themenbereiche
7	score	lid	unterrichtsstunden	prüfung	integrationskurs	xpert	deutschkenntnisse	orientierungskurs
8	prob	möglich	müssen	durchgeführt	ermäßigung	migration	kursgebühr	zusammenarbeit
8	frex	ermäßigung	migration	durchgeführt	möglich	zusammenarbeit	leider	grundlage
8	lift	migration	ermäßigung	erster	zusammenarbeit	leider	durchgeführt	wenden
8	score	migration	möglich	ermäßigung	zusammenarbeit	durchgeführt	müssen	kursgebühr
9	prob	sowie	ersten	grundlagen	teil	einführung	stehen	erhalten
9	frex	ersten	liegt	praxis	vorgestellt	gegeben	einblick	freien
9	lift	gegeben	kurstag	ersten	freien	theorie	setzt	theoretischen
9	score	gegeben	ersten	grundlagen	praxis	einführung	teil	vorgestellt
10	prob	lehrbuch	sprechen	isbn	a1	grammatik	verlag	a2
10	frex	lektion	klett	grammatik	sprachen	lehrbuch	lehrwerk	con
10	lift	sprachen	lektion	gusto	con	cornelsen	restaurant	verben
10	score	sprachen	isbn	lektion	verlag	a1	lehrbuch	klett
11	prob	eigene	bitte_mitbringen	gemeinsam	lernen	lassen	verschiedene	mitbringen
11	frex	küche	materialkosten	entstehen	kursleiterin	material	nähmaschine	kleine
11	lift	ii	schürze	zutaten	gemüse	verbrauch	abgerechnet	behälter
11	score	ii	küche	nähmaschine	materialkosten	zutaten	nähen	schürze
12	prob	einfache	schreiben	ende	verstehen	stufe	lesen	texte
12	frex	schreiben	kurze	lesen	verstehen	einfache	ende	sätze
12	lift	entspricht	vertraute	sätze	kurze	wörter	wesentliche	schreiben
12	score	entspricht	einfache	schreiben	stufe	verstehen	vertraute	ende
13	prob	innen	möchten	teilnehmer	sprache	bereits	vorkenntnisse	lernen
13	frex	möchten	bereits	erweitern	teilnehmende	innen	atmosphäre	herzlich
13	lift	geringen	auffrischen	englischen	gedacht	gelernt	kleingruppe	herzlich
13	score	geringen	innen	teilnehmer	sprache	möchten	vorkenntnissen	erweitern
14	prob	lernen	mehr	menschen	kinder	seminar	eigenen	dabei
14	frex	erkennen	seminar	besser	kinder	eltern	stimme	kindern
14	lift	referenzrahmen	körpersprache	stimme	auftreten	erkennen	konflikte	angst
14	score	referenzrahmen	seminar	kinder	stimme	erkennen	kommunikation	methoden
15	prob	vortrag	welt	jahren	seit	veranstaltung	gibt	heute
15	frex	vortrag	heute	wurde	jahren	warum	veranstaltung	welt
15	lift	südost	professor	livestream	ausgezeichnet	jahrhundert	europa	vortrag
15	score	südost	vortrag	livestream	heute	professor	wurde	jahren
16	prob	möglichkeit	kamera	digitalen	vhs.cloud	teilnehmenden	bieten	zeichnen
16	frex	digitalen	zeichnen	licht	kamera	absprache	kurs_findet	vhs.cloud
16	lift	unterrichtsmaterialien	weitergeführt	schatten	kursleitenden	präsenzunterricht	absprache	licht
16	score	unterrichtsmaterialien	digitalen	zeichnen	kamera	vhs.cloud	digitale	licht
17	prob	zeit	beim	außerdem	sowie	neben	ab	bekannt
17	frex	bekannt	außerdem	zeit	zusätzlich	kurz	wenig	jeweiligen
17	lift	bekannt	täglich	zusätzlich	kurz	jeweiligen	angesprochen	außerdem
17	score	bekannt	zeit	außerdem	neben	zusätzlich	kurz	beim

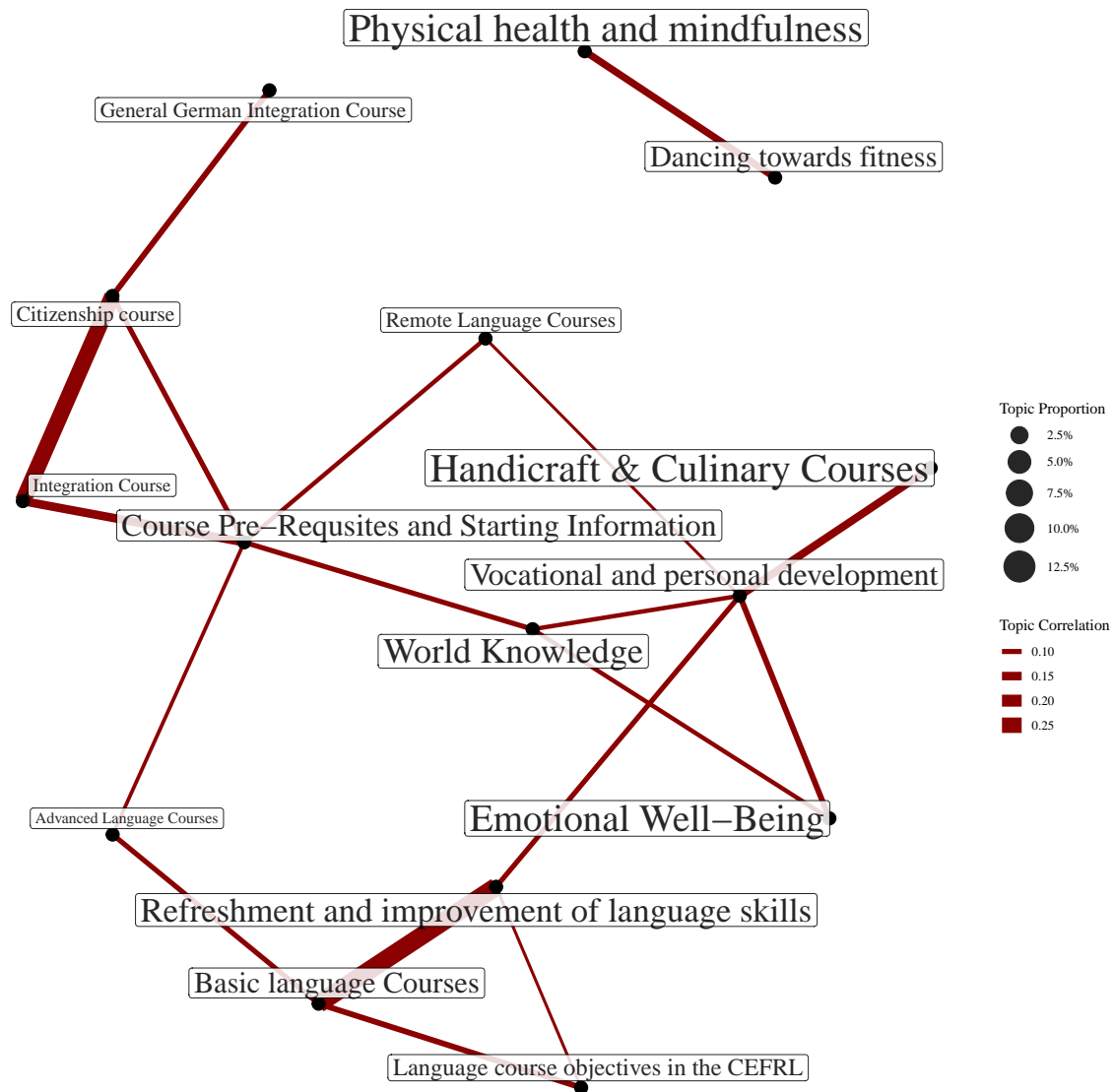


Figure 6: Correlation Network of the 16 topic model, cutoff = 0.05

Table 5: Prototypical Contexts for the Topic/Subcategory Term pairs in Figure 3

Term	Context
Course Pre-Requisites and Starting Information - Daten	zu Hause - aktuelles Smartphone mit mobilen (Vertrag oder Prepaid) -
Course Pre-Requisites and Starting Information - Daten	Teilnehmerbetreuung speichert die vhs intern persönliche - diese unterliegen selbstverständlich dem Datenschutz
Course Pre-Requisites and Starting Information - Bilanz	Gewinnermittlungsarten - Aufbau und Inhalte einer - Inventurarten und - verfahren -
Course Pre-Requisites and Starting Information - Bilanz	- Ausweis des Kapitals in der - Buchen und Auflösen von Rückstellungen
Basic and specialised commercial courses / accounting - Bilanz) - Inventur , Inventar und - Wertveränderungen in der Bilanz und
Basic and specialised commercial courses / accounting - Bilanz	Buchführung - Inventur , Inventar und - Buchen auf Bestands- und Erfolgskonten
Sector-specific specialised courses - Buchführung	zum Einsatz der Datenverarbeitung in der . Der Kursleiter empfiehlt ein Lehrbuch
Sector-specific specialised courses - Buchführung	Geschäfte : Kaufen und VerkaufenFinanzen : und GuV , Eigen- und FremdkapitalDaten
Basic and specialised commercial courses / accounting - Buchführung	Bedeutung der Buchführung - Grundsätze ordnungsgemäßer - Inventur , Inventar und Bilanz
Basic and specialised commercial courses / accounting - Buchführung	: - Notwendigkeit und Bedeutung der - Grundsätze ordnungsgemäßer Buchführung - Inventur
Basic and specialised commercial courses / accounting - Management	Erfolg sind die Analyse und das von Kosten im Hinblick auf Wirtschaftlichkeit
Basic and specialised commercial courses / accounting - Management	für Führungspositionen im mittleren und höheren . Ab 2021 bietet die VHS