Felix Terhalle

Measurement of Content Diversity in Personalized Radio

Masterarbeit

Themensteller: Prof. Dr. Detlef Schoder

Vorgelegt in der Masterprüfung im Studiengang Information Systems

der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln.

Köln, August 2021

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Consideration of diversity on the radio and in the public media is of great importance in that it is crucial for democratic discourse in society (Moore & Tambini, 2018). Diversity is seen as an “indicator of the quality of democracy” (Van Hoof et al., 2014, p. 669) and in order to achieve an appropriate diversity of opinions and topics that serve the goals of a democratic society, it is therefore important to assure diversity (Farchy, 2008). By creating diversity, citizens are not only well informed and better able to articulate their opinions, but also avoid extremes where, for example, one opinion dominates (Moore & Tambini, 2018).

Public radio stations are also obliged to play out diverse and not monotonous opinions and topics (Hirschmeier & Beule, 2018). This is recognized by the German Public Broadcasting Act which is demanding diversity in various forms as a prerequisite for the composition of the content of radio (Hirschmeier & Beule, 2018). Furthermore, the Public Broadcasting Act in Germany "explicitly states the public service remit of public radio broadcasters" (Hirschmeier & Beule, 2018, p. 3). This already indicates requirements, such as a diversity of opinions, which the public broadcasters must take into account (Hirschmeier & Beule, 2018). For this reason, it is necessary to quantify the diversity of a personalized radio stream to be able to determine the compliance. However, currently diversity is not quantified for personalized radio, new methods than those known from linear radio must be developed (Hirschmeier & Schoder, 2020).

In linear radio, program managers or other executives put the streams together (Eastman & Ferguson, 2013). But especially in individual, algorithm-controlled radio streams - often referred to as recommender system (Haim et al., 2018) - which cannot be managed by trained people, it is necessary to be able to perform automated diversity measurements. If the listener is provided with individual radio streams with a recommendation system, there is a risk that only the most important interests of the listener are displayed and only the content that is interesting for him/her is retrieved (Kunaver & Požrl, 2017). It is likely that the user is trapped in a "filter bubble" or an "echo chamber" and is only recommended content that he or she likes or that is shared in their peer group (Pariser, 2011; Sunstein, 2009). These constructs, in turn, carry the

risk of a feedback circle of self-reinforcement for the user as well as a divided society in which citizens only interact with those who think similarly (Haim et al., 2018; Levendusky, 2013; Stroud, 2008). This works against the concept of pluralism and diversity and limits citizens' decision-making options (Haim et al., 2018), since one needs a variety of viewpoints to make good decisions (Bagashka, 2014).

If the recommender system already has a mechanism that creates diversity, it is also important to control the mechanism by measuring diversity. The recommender system can then be aligned with the parameter(s) of diversity so that the accuracy of the recommender system and the diversity are both maximal (Kunaver & Požrl, 2017).

From the need to measure diversity in personalized radio, the following question is derived for the work: "How can content diversity be measured and quantified for personalized radio broadcasts?". In this way, it contributes to the measurement of diversity with the specific requirements of the personalized medium "radio". As such, the work is part of the research agenda of communication research and journalism studies. Additionally, the work is a part of the path to individualized radio. Radio should comply with state norms and represent an appropriate degree of diversity despite the personalization.

1.2 Goal of the Work

This work derives an approach to measure diversity in individualized radio as well as a visual representation for the monitoring of diversity. To achieve this goal, I first want to find out how diversity is already measured in other media and whether an approach to measuring diversity in radio already exists. Based on these findings, a measurement for the personalized radio will be developed and subsequently a possibility for the use at the public radio broadcaster Deutschlandfunk will be shown.

1.3 Organization of the Thesis

First, the fundamentals of diversity in the media and the medium of radio as well as other fundamental areas that form the basis for this work are defined. Afterwards, the context of Deutschlandfunk is described shortly. Subsequently, the scientific approach to the research question is described and already existing diversity measures from the literature are presented in a structured way. In addition, requirements for the diversity metric are derived from the Public Broadcasting Act and the literature. From these, suitable diversity parameters and a visual representation for it is derived and evaluated with Deutschlandfunk. A way is shown how

Deutschlandfunk can implement the diversity metric and the developed metric is applied to a test data set. In the final part, the limitations of the developed diversity measure are examined and whether the goal of the work has been achieved. Furthermore, the most important results of the work are summarized.

2 Theoretical Background

2.1 Media

"Media can be understood as communicative tools" (Elleström, 2021, p. 4). It can be characterized by content, such as writing, audio, video, or even art, as well as by its function to aggregate content and communicate it. The term media thus refers to the content created, as well as the companies, journalists, platforms etc. that aggregate and disseminate this content. The two do not have to take place separately (Küng et al., 2009; Vogel & Arnold, 2012).

The term media is often used synonymously with the term mass media (Dittmar, 2011), which is defined by Luhmann, 2000, p. 2 as: "[..] no interaction among those co-present can take place between sender and receivers". However, this definition excludes the notion of new media, which take place digitally and interactively (Lister, 2009). New media is characterized by the fact that the content is created by the users themselves (e.g. blogs, wikis, social networks) (Lister, 2009; Zeng et al., 2010). Since research in new media is also related to filter bubbles, echo chambers and the resulting diversity (e.g. Merry, 2016; Vaccari et al., 2016), I will also include these in the literature research, where available.

2.2 Radio

Radio is a medium (Lasar, 2016) and in this work, it is understood in the sense of broadcasting and not as a playback device or radio wave technology (Hirschmeier et al., 2020).

Different forms of radio exist from different contexts and with different program formats (Hirschmeier et al., 2020). The definition of the type of radio also refers to that of Hirschmeier et al., 2020, p. 4:

"'radio' refers to journalistic radio with a mixture of spoken-word content and music, with the journalistic spoken-word content accounting for the majority of content elements."

In contrast to other media, this radio has special characteristics. Among them, for example, the "radio hour clock", which specifies that each radio show is usually divided once again into different audio formats such as music, news, etc. (Hirschmeier et al., 2019).

In 1990, at the beginning of the Internet age, radio began to change (Lasar, 2016). New forms of radio have developed and there has been a personalization of the radio stream, which has asynchronized the respective radio streams (Lasar, 2016). Listeners receive automatically filtered streams tailored to their interests (Carlson, 2007). These systems are also called recommender algorithms/systems and create personalized streams based on listener communicated preferences (Haim et al., 2018) such as skipping posts. Some recommender systems also try to diversify the personalized stream, for example to play different topics (Toraman & Can, 2015).

There is a market division into private and public radio (Hirschmeier et al., 2019). The metric created in this work is intended to cover both forms of radio in application if they meet the characteristics described above but is intended to meet the requirements of the Public Broadcasting Act.

I use the terms personalized and individualized radio interchangeably (Hirschmeier & Schoder, 2020).

2.3 Diversity

2.3.1 Definition

Diversity is a concept used in a variety of sciences. These include ecology, psychology, sociology, economics and information science (Mcdonald & Dimmick, 2003; Stirling, 2007). Media diversity is usually used synonymously with media pluralism, as the two terms are closely related. Media pluralism refers more to social value and diversity to heterogeneity (of content, outlets, ownership etc.) (Karppinen, 2007). It is used in many empirical studies without a clear definition and rather used as a buzzword (Raeijmaekers & Maeseele, 2015) since it is difficult to define and quantify (Einstein, 2004; Entman & Wildman, 1992; D. H. Kim & Kwak, 2017). Moreover, diversity is a multidimensional problem, so mostly individual components of it are defined (Karppinen, 2006; Loecherbach et al., 2020). In computationally oriented work, diversity is also defined in terms of the method of its measurement (Loecherbach et al., 2020). If a definition of diversity is given, it usually refers to one of the most cited articles in the field, such as Napoli, 1999, McQuail, 1992, or Voakes

et al., 1996, which also define diversity by their contexts (components) (Loecherbach et al., 2020). The concept of diversity of McQuail, 1992 is often used slightly modified (Hellman, 2001; Karppinen, 2007). In his conceptualization, diversity is achieved when differences in society are represented, a wide range of choices exist, and different opinions are represented (Karppinen, 2007; McQuail, 1992). The various dimensions and contexts of diversity are discussed in the following subchapters. In summary, various contexts such as sources used, actors cited, or the actual content are examined for dispersion and difference under the term diversity.

Since diversity is a multidimensional problem, it can also be divided in empirical-quantitative and normative-qualitative aspects that are important for its description (Aslama et al., 2004). As the normative-qualitative aspects are not separable from quantification (Helberger et al., 2018; Loecherbach et al., 2020), which is the focus of this work, the normative-qualitative aspects will be described in one of the following subchapters.

2.3.2 Normative Framework

For diversity measurement, various contexts and dimensions are examined for differences. It is therefore necessary to define what these differences look like (Loecherbach et al., 2020), because the result is strongly dependent on how these normative differences are defined (Helberger et al., 2018). Decisive for the framework is the underlying democratic idea and the goal. The following table presents the four normative frameworks. It is taken from Loecherbach et al., 2020 and is derived from Bozdag & van den Hoven, 2015; Helberger, 2015; Möller et al., 2018; Raeijmaekers & Maeseele, 2015 who describe the same four frameworks.

Normative Framework	**Focus**	**Main Goal**
Liberal-aggregative	Market	Reflection of social heterogeneity (reflective diversity), Mirroring society
Liberal-individual	Consumer	Autonomy of individual, Consumer satisfaction, No restriction to choose
Deliberative	Public Sphere	Inclusive public debate, All possible perspectives, Equal share of viewpoints etc. (open diversity)
Adversarial	Alternative Voices	Promoting minorities, Share of minority voices favored

Table 1: Normative Frameworks

The liberal-aggregative framework wants the media to reflect societal heterogeneity in that ideas and opinions compete against each other in the so-called "marketplace of ideas" (Entman & Wildman, 1992). The liberal-individual framework, which is also known as the marketplace model describes the view from the consumer's perspective. Diversity is achieved by covering all consumer interests (Aslama et al., 2004; Hellman, 2001). Moreover, more choice equals more diversity (Hellman, 2001). These contrast with the Public Policy Model (deliberative framework), which sees diversity as a normative criterion of quality. The task is not to meet demand, but pluralism at different levels as well as equal access for all population groups (Hellman, 2001; McQuail, 1992). The goal is to reflect all entities, sources and topics with the same proportions (Loecherbach et al., 2020). The fourth framework (adversarial) assumes that minorities and small groups in democracies are often at a disadvantage due to structural problems and therefore need to be supported by a focus in the media (Bozdag & van den Hoven, 2015).

To capture diversity as an actual state in this work and to measure what is played out in the individual radio stream, this work considers the deliberative framework. The focus is on measuring the diversity specified in the Public Broadcasting Act. However, the requirements of the Public Broadcasting Act are only vaguely formulated (e.g., "[..] service offer should be balanced (§11 (2))." (Hirschmeier & Beule, 2018, p. 3)). That is why a framework that assumes equal distribution and does not consider target values is assumed in this work. In the

literature search for diversity metrics, all normative frameworks are considered since some metrics are applicable to various frameworks.

2.3.3 Open vs. Reflective

Another distinguishing feature of diversity is reflective and open diversity (McQuail, 1992). Open diversity is based on the views of the deliberative normative framework (Loecherbach et al., 2020). "Open diversity exists when content is as heterogeneous as possible" (Van der Wurff, 2004, p. 217). The content should be as equally distributed as possible - across opinions, categories, etc. (McQuail, 1992; Van der Wurff, 2004). Reflective diversity deals with the divergence between supply and demand (McQuail, 1992; Van der Wurff, 2004) and is based on the liberal-aggregative normative framework (Loecherbach et al., 2020). Content should reflect society outside the media landscape as closely as possible (Van Hoof et al., 2014).

Due to the fact that the Public Broadcasting Act only states that the program must be balanced (Hirschmeier & Beule, 2018) and no target values are specified, equal distribution is assumed in this work. Moreover, with open diversity it is possible to represent the as-is state. In this paper, following the deliberative framework, diversity is thus considered in the context of open diversity.

2.3.4 Contexts

Diversity can be seen from different contexts (Karppinen, 2007; Mcdonald & Dimmick, 2003). It is measured and viewed from different directions in each case. The following paragraphs will highlight each of these contexts.

The exposure diversity context refers to the diversity that a user actually experiences through their own selection of media content (McQuail, 1992). It therefore does not refer to the media available to the user, but only to the sources and content that the user actually consumes (Napoli, 2011).

In the source diversity context, the sources of the content are examined. It is also called structural diversity (Napoli, 1999). It “includes everything concerned with how the content was made and by whom” (Loecherbach et al., 2020, p. 612). For example, the number of competitors in the market (Napoli, 1999).

The content diversity context is divided into three sub-contexts (Napoli, 1999). The first is viewpoint diversity, which “refers to the diversity of viewpoints and of social, political, and cultural perspectives represented within the media” (Napoli, 1999, p. 22). For example, different opinions on a topic would mean a greater

viewpoint diversity. The second context is program diversity. This refers to diversity that comes from different types of programming, such as news, music, or different types of shows (McQuail, 1992; Napoli, 1999). Loecherbach et al., 2020 refer to it as topic diversity, since this covers not only the range of program type, but also the topics. The third context describes how the population is represented in the content of the medium. For example, whether minorities or other population groups are represented and get to speak. Again, there exists a somewhat broad term that is often used in the literature: Diversity of entities (Loecherbach et al., 2020). In addition to population groups, this can also be applied more generally, such as to organizations.

Further contexts are the perceived diversity and the audience diversity. The perceived diversity which is based on the user's perception of diversity (Hoffman et al., 2015) should be distinguished from exposure diversity, which does not examine the perception of diversity, but rather the selection of available content. E.g., men and women might perceive diversity differently (Ahern, 2011). Audience diversity refers to the nature of the audience, such as demographics or regularity of use of a medium (McQuail, 1992).

2.3.5 Exposure vs. Supply

The supply side or opposite side of exposure diversity is the supply diversity. This is the dominant side in science and can be identified by analyzing the media content (Loecherbach et al., 2020). Supply diversity consists of source diversity as well as content diversity described in 2.3.4 Contexts. It is also called internal diversity (Moore & Tambini, 2018).

In the traditional view of the main contexts of diversity (see Figure 1 adapted from Napoli, 2011), it is assumed that source diversity affects content diversity (Napoli, 2011). However, there is no final proof of this causal relationship (Napoli, 2011). It is rather questioned whether this connection really exists (Horwitz, 2005). Additionally, the focus from source and content diversity currently shifts to content and exposure diversity (Loecherbach et al., 2020; Napoli, 2011).

Figure 1: Main Contexts of Diversity

In this work, the supply side with a content diversity aspect is considered. This is the diversity that can be measured from the perspective of a radio provider.

2.3.6 Dimensions

Diversity can be simply represented mathematically as quantitative measure. As Junge, 1994 puts it:

"In statistical terms a measure (index) of diversity is a summary description of a population with a class structure. More generally, quantification of diversity is related to the apportionment of some quantity (e.g., number of elements, time, mass) into a number of well-defined classes." (Junge, 1994, p. 16).

It follows that diversity cannot be measured until the elements to be measured are divided into classes. This can be, for example, the categorization of articles by resort in a journal. Stirling, 2007 considers this in a rather larger context and speaks of "apportioning of elements or options in any system" (Stirling, 2007, p. 709).

To measure the diversity of the contexts indicated above, diversity must be made quantifiable for these. For this reason, Stirling, 1994 already attempted to ascribe dimensions to diversity in the field of electricity supply. In 1994, he described the dimensions variety, balance, and disparity. These were used in a variety of research fields and until 2007, no other dimension could be found until then either (Stirling, 2007). Stirling, 2007 also shows that all three dimensions are already quantifiable. Dimensions are ways to measure diversity in different ways, with elements (e.g., articles) placed in classes/systems.

The first is variety, it describes a single metric to measure diversity. It is the answer to the question "how many types of thing[s] do we have?" (Stirling, 2007, p. 709), among others, how many categories the elements were divided into.

The second is balance, which describes the distribution of elements among categories (Loecherbach et al., 2020). It is the answer to the question "how much of

each type of thing[s] do we have?" (Stirling, 2007, p. 709). It is also called evenness or concentration measure (Stirling, 2007).

The third is disparity, it describes the distance between elements (Loecherbach et al., 2020; Stirling, 2007) and is the answer to the question "how different from each other are the types of thing that we have?" (Stirling, 2007, p. 709).

These dimensions of diversity can be measured either alone or in combination. Mcdonald & Dimmick, 2003 use the term dual-concept diversity in the context of television and radio, which combines the dimensions variety and balance and thus reflects the number of classes and the distribution of elements among these classes. A one-dimensional metric that measures only one of the dimensions of diversity is called single-concept diversity. Usually this is the measurement of variety, for example of the sources or of the content (Mcdonald & Dimmick, 2003).

Moreover, Mcdonald & Dimmick, 2003 suggest that it is not sufficient to measure only one of the dimensions. Suppose you have five classes where the first one accounts for 80% of the elements and the rest of the classes account for 5% each. This already appears to be non-diverse when looking at the distribution (see Figure 2) and is also not complete according to the mathematical definition of Junge, 1994, according to which diversity is measured as a "[..] summary description of a population with a class structure" (Junge, 1994, p. 16). Even if only the balance and, for example, the number of classes (variety) is not considered, one of the distributions may have more classes than the other and the balance would remain the same (Mcdonald & Dimmick, 2003). The same is true for disparity. Regardless of how many classes there are or how they are distributed, the classes can be very close or very far from each other thematically. Thus, to get a comprehensive picture of diversity in most cases, it is not enough to measure only one of the dimensions. Hence, all dimensions are considered in this work.

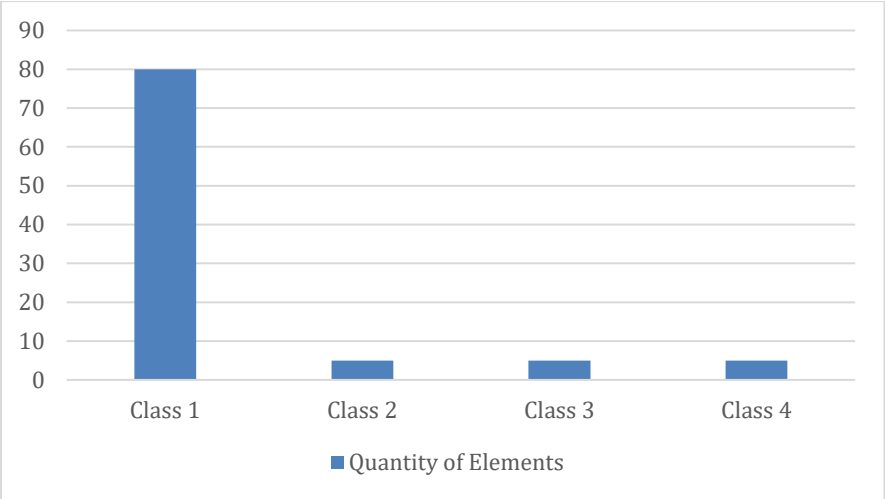

Figure 2: Single Concept Diversity Distribution

2.3.7 Vertical vs. Horizontal

For each of the diversity dimensions it is possible to examine by vertical - also called channel diversity (Aslama et al., 2004) - and/or horizontal view. The former denotes the examination over a period of time of a single channel. “It explains how programs were concentrated into a limited number of genres or how they were distributed over many genres” (Einstein, 2004, p. 5). The latter denotes the examination over a period of time of multiple channels (Entman & Wildman, 1992; Hellman, 2001; McQuail, 1992; Napoli, 1999). The investigation of an individual radio stream of Deutschlandfunk as it is done in this work is a vertical investigation since each stream must be measured individually.

2.4 Public Broadcasting Act

The Public Broadcasting Act is a set of requirements to which public broadcasters must comply (Hirschmeier & Schoder, 2020). It represents requirements that promote free speech and thus support a democratic system overall. Broadcasters must comply with these requirements, and their compliance must be reported. For example, Deutschlandfunk has to report every two years on the fulfillment of their mandate and on the quality and quantity of its offerings (German Federal States, 2020). Broadcasters do this in exchange for public funding (Hirschmeier & Schoder, 2020).

3 Deutschlandfunk

3.1 General

The name Deutschlandfunk belongs to Deutschlandradio and refers to a radio station. The company Deutschlandradio is a corporation under public law and thus performs tasks for the German state (Bundeszentrale für politische Bildung, 2016). It is stipulated that an objective overview of world events should be provided and human dignity is to be respected and protected, as well as the cultural diversity of the countries is to be taken into account (Deutschlandradio, 2019). Furthermore, and most important, the requirements §1 to §15 of the Public Broadcasting Act are also declared mandatory in the Deutschlandradio contract (Deutschlandradio, 2019).

Deutschlandfunk operates throughout Germany and broadcasts politics, business, science, culture, and music (Deutschlandradio, 2021b). According to the station's own information, the proportion of words is 80% (Deutschlandradio, 2021a). It is split into the stations Deutschlandfunk, Deutschlandfunk Kultur and Deutschlandfunk Nova (Deutschlandradio, 2021d).

3.2 Individual Radio Stream

In cooperation with the University of Cologne and the Department of Information Systems and Information Management, Deutschlandfunk is researching the possibilities of individual radio and, as a result, can provide useful information on the topic of personalized radio as well as insights into the business of radio making. They can also provide feedback from different roles and perspectives from their organization.

4 Methodology

To develop a diversity measure and a corresponding visualization suitable for radio, a mixed method approach is conducted. First, a literature research on the topic of diversity measurement in the media is conducted. An explicit search for existing diversity measurements in radio is also realized. For this purpose, a keyword search is used in various literature search engines as well as in media relevant for radio, including German language media. In a more detailed analysis of the results obtained from the literature search, the individual approaches to measuring diversity are classified and inductively subdivided into a matrix according to concepts. This is done

by using Webster and Watson's approach for literature reviews as well as the "concept matrix" approach, which is used to group similar concepts found in the literature (Webster & Watson, 2002). Following the literature review, requirements for the diversity metric and its graphical representation such as requirements of the Public Broadcasting Act or specific characteristics of the individual radio stream are derived. The requirements and the concept matrix are then used to derive an approach for measuring content diversity in personalized radio. Second, a graphical monitoring mockup for the developed diversity measure is created. Interviews with various experts from Deutschlandfunk are conducted and the measure and graphical interface are then adapted. Here, the focus is on the design and subsequent evaluation of the dashboard using a design science approach.

The following chapters describe the literature research in detail and give reasons why the transfer of diversity metrics from media to radio is possible.

4.1 Literature Research

First, the literature review is conducted on the topic of diversity measurement in the media. Since an initial literature search, the very recent review by Loecherbach et al., 2020 was found. It comprehensively addresses the topic of diversity in the media and presents an evaluation that already classifies the literature that measures diversity into variety, balance, and disparity (table B2 in Loecherbach et al., 2020). This is taken as the basis of the review.

But since Loecherbach et al., 2020's focus is not on the metrics, but rather on the general concept of diversity assessment (variety, balance, disparity), each paper from Loecherbach et al., 2020's evaluation was again examined for the diversity measure and put into a concept matrix as a first step. Duplicate entries (1) from Loecherbach et al., 2020 are summarized. In addition, Loecherbach et al., 2020's papers were filtered for diversity of content. E.g. exposure diversity was filtered out. Diversity measurement methods that do not make quantitative statements are left out, too. The authors provided additional internal information about the literature. Among other things, which diversity measure was used. However, only 38 of the 133 articles state a metric.

As a second step, a review complementary to Loecherbach et al., 2020 was conducted, which also included radio and television (Loecherbach et al., 2020 excluded these). In this supplementary review, Google Scholar and ProQuest were

used. Analogous to Loecherbach et al., 2020, the same starting point was chosen: Work from 1998 is included until today, 2021 (Loecherbach: 2018). This time period is intended to include methods that originate from both offline and online media (Loecherbach et al., 2020). The following queries are examples of queries used with their associated search strings. Different search strings had to be used for each search engine, as no relevant results could be obtained with the others:

- Google Scholar: (tv OR television OR radio OR broadcast) AND (measure*) AND (diversity OR pluralism OR similarity)
- ProQuest ABI/Inform Complete: (tv OR television OR radio OR broadcast) AND (measure* OR assess*) AND (diversity OR pluralism OR similarity)

Furthermore, two journals from the broadcast environment were explicitly searched ("Journal of Radio & Audio Media" and "Journal of Broadcasting & Electronic Media") as well as a German-language journal (“kommunikation@gesellschaft”), which has dealt with the topic broadcasting in the past (e.g. Pöchhacker et al., 2017).

After using search engines and direct searches in the journals, a forward and backward reference search of the literature was performed - as described in Webster & Watson, 2002. The following figure shows the process of paper selection, where n is the number of papers:

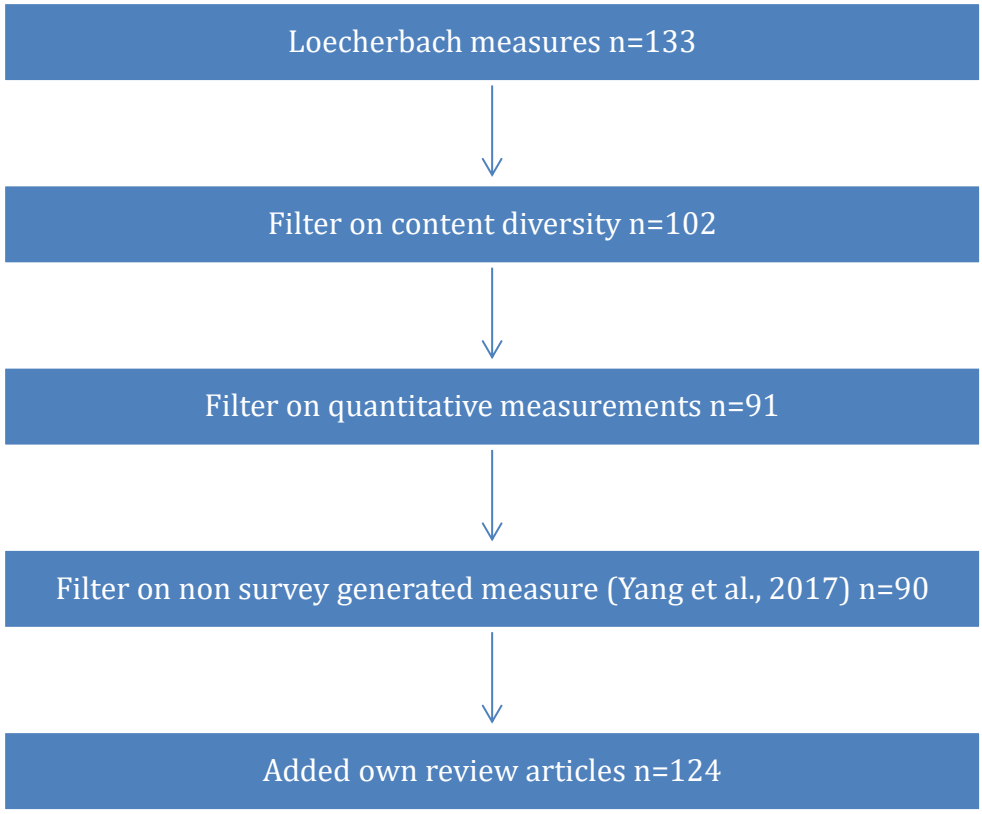

Figure 3: Selected Papers

In the supplementary review, papers were initially shortlisted if they indicated in the abstract, title, or keywords that they measure content from media in terms of its diversity. If it is noticed that relevant results are not captured when searching with the search engines, additional search queries were created. In the next step the full text of the results was checked. If the result was still relevant according to the above criteria, the approaches to measuring diversity were put into the concept matrix.

The literature search also excludes everything that does not examine the subject content. E.g. Maekawa et al., 2014 are examining the spelling - syntax - and no conclusion is drawn on the actual content or the statement of the text.

4.2 Rationale

The transfer of metrics from non-broadcast media to broadcast media such as television and radio is possible because the contributions are very similar in structure and type and differ mainly in medium. Van Leuven et al., 2015 already uses news articles from television in a common data set like articles from newspapers. Also Swert & Hooghe, 2010 as well as Zeldes et al., 2012 use similar methods for television as used for example for newspapers (building categories, counting issues/elements).

5 Diversity Measure in Literature

The papers identified by the Loecherbach et al., 2020 literature review already show a clear trend towards measuring the dimension variety. Not in the classical way that the individual categories are simply counted, but that articles are divided into categories (mostly topic or viewpoint oriented) and the proportions of the categories per outlet (or something similar) are reported. This trend continues in the extended literature review for broadcast. Another trend is evident with a strong focus on the topic context.

In part, the results from Loecherbach et al., 2020 differ at the paper level with those of this thesis because the focus in the literature review was different: content diversity. Therefore, it is possible that articles that examined various contexts of diversity are classified differently in this work. For example, a paper is classified with the topic context in Loecherbach et al., 2020 and here it is classified as viewpoint context in this paper, because for content diversity the viewpoint context was measured and for topic diversity exposure diversity.

The following chapters in 5.1 Diversity Measurements are divided by dimension and show the result of the literature review. For each of the dimensions (variety, balance, and disparity), the metrics that measure that dimension are described as another sub-chapter. Here, a metric has an own sub-chapter if a minimum of two papers used the metric. This is to ensure that no specialized or unestablished metric is adopted for the context of personalized radio. An overview (concept matrix) over all diversity metrics from the literature is given in appendix A Diversity Measurement Overview.

5.1 Diversity Measurements

5.1.1 Variety

5.1.1.1 Divide, Count, and Report

Count frames per article is a collection of methods that count various types of features based on an article or post in the media. This category also includes papers that subdivide by different frames or classes and then represent the distribution of these frames and classes without defining a "single variable" as a variety count. These types of measures as well as the category count from the next chapter are called “single-concept diversity” (Mcdonald & Dimmick, 2003).

For example, Benson, 2009 assigns pre-defined frames to each article on immigration such as "Job Threat," "Public Order Threat," or more problem-oriented frames such as "racism" or "Cultural Diversity". In addition, he assigns frames to each article based on which entities are cited in the article (ex. "Executive/Bureaucratic" or "Trade unions"). Both frame types are counted per article and the average of frames per article is used as a metric at the newspaper level to represent diversity. Also, S. Rodgers et al., 2000 count the frames in different areas for the appearing characters like age and role (entity) or frames related to whole articles like topic or political ideology (topic, viewpoint). However, S. Rodgers et al., 2000 does not average the frames per news outlet, but only reports the count per category.

5.1.1.2 Category Count

Counting categories is the most common way to measure diversity after counting frames per article. Media articles are divided into categories. These can be different political viewpoints as well as topics or occurring sources. Fernández-Quijada, 2017, for example, are guided by a list of the European Broadcasting Union, which stores information on genres and main topics. A subset of media articles can then be divided into these categories. The categories are then counted, and the count is used as a diversity metric. Unlike "Count frames per article", there is always a single metric that can be used as a diversity indicator. Chambers, 2003 uses this metric already in the context of radio and calculates how many different program types were played on average in the radio stations studied. Lacy et al., 2012 count the different source types per item and report the average source types cumulatively.

One advantage of counting categories is the simplicity of this metric, which can be understood without a mathematical formula. A disadvantage of this metric can be that oligopolies formed in the media are controlled by a few owners and form similar genres/categories to the detriment of genre/category diversity in order to gain efficiency (Chambers, 2003; Wirth, 2001).

5.1.2 Balance

5.1.2.1 Simpson Index

The Simpson index or Simpson's D was discovered in 1949 by Edward H. Simpson (Simpson, 1949) and originates from economics, where it is used to measure the concentration of the market (Boydstun et al., 2014). “It takes into account the

numbers of categories as well as the distribution of the elements across these categories" (Vergeer et al., 2012, p. 88). There are several similar measures around it, such as Simpson's D_z (Simpson's D normalized form), Junge's H or the Herfindahl-Hirschman Index (HHI), which was invented parallelly (Mcdonald & Dimmick, 2003). Simpson's D is calculated as follows (Y.-C. Lee, 2007; Simpson, 1949; Tan & Weaver, 2013):

$$D = 1 - \sum_{i=1}^{k} {p_i}^2$$

Where p_i is the proportion of the respected category i in the population. k is the total number of categories. The squared proportions are summed and subtracted from the sum of 1 (Carpenter, 2010).

It is, unlike Shannon's H simple to calculate and can be easily interpreted (Y.-C. Lee, 2007; Mcdonald & Lin, 2004). The index ranges from 0 to 1 (Mcdonald & Dimmick, 2003). It gives the probability of two randomly selected articles or posts to be of the same category (Simpson, 1949).

Simpson's D is sensitive to the number of categories or objects used. So Masini et al., 2018 use Simpson's D_z, which is more suitable for comparing different measurements with different underlying number of categories or objects (Y.-C. Lee, 2007). Still, Vergeer et al., 2012 and S. Park, 2011 claim, that Simpson's D is a sensitive indicator "to the proportion of objects allocated to each category" (S. Park, 2011, p. 522).

According to Mcdonald & Dimmick, 2003 the Herfindahl-Hirschman Index is mathematically identical as an index, but more difficult to interpret because its numerical range is not exclusively from 0 to 1. Instead, it ranges from 0 to 10,000 (Entman, 2006; S. Park, 2005). The index is not optimal for all applications because it does not have high precision near the maximum and minimum (Boydstun et al., 2014). Since its calculation is identical to Simpson's D, the same can be assumed for it. Gini also established the same formula several years before Simpson, yet the index is usually referred to as the Simpson index (Mcdonald & Dimmick, 2003).

As an example of Simpson's D, consider Masini et al., 2018, which use the D to calculate the balance of actors. They divide the actors into categories (civil society, journalists, business etc.) and calculate a normalized D of 0.82. Since D is close to 1 in this case, it can be assumed that the actor categories are balanced to some degree.

5.1.2.1.1 Multivariate

Another extension on Simpson's D is D_{multi}, the multivariate D (Mcdonald & Lin, 2004). This is used when not only classifying by one variable but using one or more others - such as political orientation and genre (Agresti & Agresti, 1978; Mcdonald & Lin, 2004). D_{multi} represents the average Simpson's D across different category systems. The following formula represents the calculation (Y.-C. Lee, 2007):

$$D_{multi} = \frac{1\left(\sum_{k=1}^{k_1} {p_i}^2 + \sum_{k=1}^{k_2} {p_i}^2 + \cdots + \sum_{k=1}^{k_m} {p_i}^2\right)}{m}$$

Where p_i is the proportion of the respected category i in the population, m is the number of different category systems and l the level of m ($l = 1,2, \ldots, m$). k_l is the total number of categories in the corresponding category system.

5.1.2.1.2 Lieberson

Lieberson's D (D_b) is a variant of the D developed by Lieberson (Lieberson, 1969). It examines two different classifications for similarity (Agresti & Agresti, 1978; Y.-C. Lee, 2007; Mcdonald & Lin, 2004). Lieberson's D represents the probability that two randomly drawn items from two different pools were classified differently (Mcdonald & Lin, 2004). It is calculated as follows:

$$D_b = 1 - \sum_{i=1}^{k} p_i q_i$$

Where p_i is the proportion of the respected category i in the first population. q_i is the proportion of the respected category i in the second population. k is the total number of categories.

This measure is particularly useful for testing two distributions for similarity in terms of classification (Mcdonald & Lin, 2004). For example, Y.-C. Lee, 2007 use Lieberson's D to compare traditional newspapers with the overall diversity of traditional newspapers and can determine how specialized individual newspapers are. And, like the multivariate D, Lieberson's D also has a multivariate version (Agresti & Agresti, 1978).

5.1.2.2 Shannon Entropy

Entropy is a measure of average information level (Shannon, 1948) stemming from information theory (Boydstun et al., 2014). It measures both the dimension variety and the balance of the data set (Hellman, 2001). It is also called H-statistic

(Culbertson, 2007; Huang, 2010). In communication science, it measures the relative distribution of attention across all objects or categories (Jonkman et al., 2018). The nature of the categories or objects can differ and can, for example, refer to entities as well as topic categories (Boydstun et al., 2014). “It is sensitive to the amount of categories and evenness but is mathematically robust regarding small samples“ (Steiner et al., 2019, p. 8). Shannon's H is one of the best-known metrics and is mathematically very similar to Simpson's D (An et al., 2014; Mcdonald & Dimmick, 2003). It is calculated by multiplying the negative sum of the distributions by the logarithm of the distribution (Mcdonald & Dimmick, 2003):

$$H_{abs} = -\sum_{i=1}^{k} p_i \log_2(p_i)$$

Where p_i is the proportion of the respected category i in the population. k is the total number of categories.

Mcdonald & Dimmick, 2003 also find that compared to the other diversity metrics used, Shannon's H as well as Simpson's D are particularly sensitive to the number of categories and the maximum proportion. An advantage over Simpson's D is that Shannon's H has higher accuracy at the maximum and minimum values (near 1 and 0) (Boydstun et al., 2014).

Since a biased result can also occur with the Shannon entropy because the sample size is not considered An et al., 2014 additionally use the Miller-Madow correction technique.

Furthermore, there are two subtypes of entropy, which are presented in the following chapters:

5.1.2.2.1 Relative Entropy

For the relative entropy (or standardized entropy), the values range between 0 and 1. 0 indicates homogeneity and 1 maximum diversity (Aslama et al., 2004; Peter & Vreese, 2003). Due to its logarithmic mathematical structure, it becomes more difficult to increase the value as it gets closer to full diversity (Aslama et al., 2004; Hellman, 2001). According to Aslama et al., 2004, relative entropy is calculated as follows:

$$H_{rel} = \frac{H_{abs}}{H_{max}}$$

where

$$H_{max} = log_2(k)$$

Shannon's H is divided by the logarithm of the number of the categories k. This normalized version should be used if the total number of categories differs markably between different measurements (Boydstun et al., 2014).

For example, Hellman, 2001 uses relative entropy to calculate vertical diversity in a category system consisting of 15 genres. He also shows that it is possible to relate relative entropy to others. He calculates the diversity of all channels occurring in his study and then subtracts one channel after another, exploring how much diversity each channel contributes to the total diversity.

5.1.2.2.2 Entropy in Number Equivalents

Entropy in number equivalents ranges from 1 to the number of categories (k) (Van Hoof et al., 2014). It represents the number of categories or objects that "receive a substantial and equivalent amount of attention" (Van Hoof et al., 2014, p. 675). It can be calculated as follows (Takens et al., 2010; Van Hoof et al., 2014):

$$H_{ne} = \sum_{i=1}^{k} (\frac{1}{p_i})^{p_i}$$

For each category, 1 is divided by the proportion of attention p_i of the respected category i and again raised to the power of p_i. k is the total number of categories. The sum of all values gives the diversity.

The advantage of this type of entropy is that it is easier to interpret, since it is very close to the number of categories (Takens et al., 2010).

Takens et al., 2010 for example calculate the entropy in number equivalents for the number of mentioned parties in a newspaper. Then they calculate the differences between different newspapers, which are also easy to interpret.

5.1.2.3 Ordinal Scale

The ordinal scale measures include classifications that can be represented on a scale. It is used to represent the balance of the classification based on the classes arranged on a scale.

Seo, 2018 use a scale from 1 to 4, on which is represented how soft or hard the used theme or framing is. Soft topics are topics that are more personal and emotional and have less public interest. With this scheme and the average values, the authors can

summarize in their paper that a particular newspaper in North Korea tends strongly toward softer topics.

Day & Golan, 2005 use a positive/negative/neutral scale to examine the evaluation of various topics such as "gay marriage" and can evaluate them quantitatively by mapping them to a scale of 1 to 3 and taking the average.

5.1.2.4 Open and Reflective Diversity

Following the notions of the terms open and reflective diversity, Van der Wurff & Van Cuilenburg, 2001 create simple diversity metrics that map these. Open diversity assumes that maximum heterogeneity is good, i.e., if all elements in the respective categories are equally distributed, open diversity is 1 (Van der Wurff, 2005). This is calculated by comparing the proportion p_i from the respective category i with the target proportion $\frac{1}{k}$ of each category, where k is the total category count (Van der Wurff, 2004). Open diversity is calculated as follows:

$$Open\ diversity = 1 - \sum_{i=1}^{k} \frac{p_i - \frac{1}{k}}{2}$$

Reflective diversity, on the other hand, assumes that the proportions of the categories must be different and compares the proportions p_i with pre-defined target values d_i for the respective category i (Van der Wurff & Van Cuilenburg, 2001). k is the total category count. It is 1 if the target values match the actual values (Van der Wurff, 2005):

$$Reflective\ diversity = 1 - \sum_{i=1}^{k} \frac{p_i - d_i}{2}$$

A very similar approach is taken by Hughes & Prado, 2015, who calculate the relative deviation of stakeholder group distribution compared to those from the population.

The main advantage in the formulas is that they follow and map the definitions of open and reflective diversity (Van der Wurff, 2004). Moreover, this metric is also easy to understand and interpret.

5.1.2.5 Deviation Index

Similar to the metric of Van der Wurff & Van Cuilenburg, 2001 on reflective diversity, the Deviation index measures the deviation not from a given distribution, but from two different programs (Hellman, 2001; Powers & Benson, 2014). With this,

it is possible, for example, to define one program as a model and compare to it. It is also called “mean absolute deviation” (Benson et al., 2018). Hellman, 2001 compares the number of minutes played per category by two different broadcasters. The deviation is the sum of how many minutes all categories together deviate from each other.

5.1.2.6 Remaining Measures

Other diversity metrics that appear only once in the researched dataset include the index of qualitative variation, which measures how equally items are divided into categories (Woods, 2007). A simple metric that assumes items are balanced if they represent more than one viewpoint (S. Rodgers et al., 2000). And a residual analysis by Humanes, 2013, which examines how much importance is assigned to individual topics or categories.

5.1.3 Disparity

5.1.3.1 Chi-Square Test

The general idea behind the Chi-Square test is to use "frequencies of co-occurring or common words to calculate similarities between documents" (Ibrahimov et al., 2002, p. 296). The Chi-Square test can be used to test two distributions for similarity. For example, Ibrahimov et al., 2002 compared the similarity of documents using word distributions. For this purpose, they used the following formula of the Chi-Square test as a basis:

$$X^2 = \sum_{i=1}^{n} \frac{(observed\ frequency - expected\ frequency)^2}{expected\ frequency}$$

As observed and expected distributions the weight of a word from document i and document j is inserted. The weight of a word is based on its frequency. The result is a similarity over all words n of both documents.

The Chi-square similarity function calculates more accurate values than the Cosine similarity function especially for spoken-word documents (Ibrahimov et al., 2002).

5.1.3.2 Cosine Similarity

The solution approach for a similarity measure is the same for Cosine similarity as for the Chi-Square test, namely measuring distances of words in documents

(Ibrahimov et al., 2002). But where the Chi-Square test uses the distributions to calculate similarity, the Cosine similarity calculates the angle between two vectors in vector space (Welbers et al., 2016). The value range is from 0 to 1, where 1 indicates an identical document (Ibrahimov et al., 2002; Welbers et al., 2016). It is calculated with the following formula (Ibrahimov et al., 2002; Yu & Zhou, 2009):

$$similarity(d_i, d_j) = \frac{\sum_{k=1}^{n} w_{ki} w_{kj}}{\sqrt{\sum_{k=1}^{n} w_{ki}^2 \sum_{k=1}^{n} w_{kj}^2}}$$

Where d_i, d_j represent the documents to be compared and w_{ki}, w_{kj}, the respective word vector components for the documents for word k. n is the maximum number of vocabularies in vector space.

L. Li et al., 2014 e.g., use Cosine similarity to compare the similarity of document sets suggested by different recommender algorithms. In doing so, they assume that the lowest similarity is likely to have the highest diversity.

5.1.3.3 Jaccard Index

The Jaccard index uses the frequency and dependencies of words to calculate a similarity of documents (Ibrahimov et al., 2002). It was invented in 1901 and is a "well-known measurement of the similarity between two sets" (Fletcher & Islam, 2018, p. 4). It is calculated by dividing the intersection of both sets by the union of both (Fletcher & Islam, 2018).

The Jaccard index gives results similar in accuracy to the Chi-Square test and better than the Cosine similarity function (Ibrahimov et al., 2002).

Similar to L. Li et al., 2014, Zafar et al., 2015 use the Jaccard index to measure the similarity of identified sets of topics to find similar ones.

5.1.3.4 Remaining Measures

Other diversity metrics that cover disparity include a uniqueness percentage that indicates how unique the set of songs is (Martin, 2004) and the sum of distances to variety index that looks at the distance between all program category pairs (Farchy & Ranaivoson, 2011). The Jenson-Shannon Divergence, which is used by González-Bailón & Paltoglou, 2015 is similar to the relative entropy but calculates the distance between topics.

5.1.4 Stirling Index

The Stirling index is a special case. It measures all three dimensions and is applied only once in the papers of the research (Farchy & Ranaivoson, 2011). Nevertheless, I chose to describe it, as it could provide good incentives for how to unify individual diversity measures from different dimensions.

Stirling, 2007 created a generalized diversity index independent of the research field because he noticed that the requirements for diversity are very similar. The Stirling index is a measure of diversity that covers all three dimensions of diversity (Farchy & Ranaivoson, 2011). It is motivated by different measurements for variety, balance and disparity (Stirling, 2007). The basic form ("balance/disparity-weighted variety") is calculated as follows:

$$D = \sum_{ij\,(i \neq j)} d_{ij} p_i p_j$$

The formula sums over the half matrix of $\frac{n^2-n}{2}$ where n is the number of documents. For each element in the matrix, the disparity between document i and j is multiplied by the balance of i and j, where p_i and p_j indicate the relative proportions of all documents. d_{ij} is the disparity between document i and j.

Although the formula has already been applied in the media context (Farchy & Ranaivoson, 2011; Moreau & Peltier, 2004), it leaves open how disparity is calculated. Stirling, 2007 explains this by the fact that disparity can be very different depending on the application. For example, it can be a real number to be calculated or assigned to one of two categories (Stirling, 2007).

5.2 Scaling of Diversity Measure

The measured diversity refers in each case to a set of elements. Since the size of the elements varies (e.g., article length in words, post length in minutes). This imbalance must be considered (Champion, 2015; Hellman, 2001). For newspaper articles, for example, this is done with the article or paragraph length (Culbertson, 2007). But in broadcasting, the different article lengths are also taken into account. Scott et al., 2010 offset the category count with post length in seconds: "each broadcast had a cumulative total length in seconds. This allowed for a pooled total (in seconds) of each variable category per broadcast" (Scott et al., 2010, p. 330). Also, Peter & Vreese, 2003 do this for broadcasting: "we based our computation on the overall length of the coverage of a particular issue category during the period of investigation rather

than on the overall number of stories" (Peter & Vreese, 2003, p. 52). Mangani & Tarrini, 2018 even take time as the actual measure for entity diversity by taking the "speaking time" of an entity (e.g., a person) as a measure.

5.3 Omission of Genres

In broadcasting, certain issue categories are not always considered relevant for the diversity calculation. Therefore, these are omitted. Humanes, 2013, for example, exclude weather and sport features on television.

6 Development of Diversity Measure

6.1 Requirements

Requirements for diversity metrics for personalized radio are collected below. The requirements do justice to both the public mandate to promote democracy and the specific requirements of the medium of (personalized) radio. They are separated by origin of the requirement.

6.1.1 Public Broadcasting Act

As a basis for a diversity metric, this work includes Hirschmeier & Beule, 2018's summary of the requirements of the Public Broadcasting Act (p. 3). These have remained the same even with the introduction of the new Public Broadcasting Act in 2020 (Federal States of Germany, 2020). They are particularly relevant because they contain mandatory regulations for public radio stations (including Deutschlandfunk). From these, requirements for the diversity metric can be derived:

R1. "The service offer of public broadcasters should include information, education, culture, and entertainment (§ 11 (1))."": The diversity metric should indicate if all categories are played (variety).

R2. "The service offer should give a comprehensive overview over international, national, and regional events in all essential areas of life (§ 11 (1))."": The diversity metric should indicate that international, regional, and national events as well as "essential areas of life" are covered.

R3. "The service offer should reflect the diversity of opinions (§ 11 (2))." and "The service offer should support the process of forming a free and

individual opinion and therefore fulfil the needs of a democratic, social and cultural society (§ 11f (4)).": The diversity metric should indicate how diverse the opinions of the posts played out are.

R4. "The service offer should be balanced (§ 11 (2)).": All the above and following requirements should be balanced according to their categories.

Other European countries have similar requirements for public service media, which are already covered by the listed requirements above (Iosifidis, 2014).

6.1.2 Literature review

In addition to the requirements from Hirschmeier & Beule, 2018, the following requirements are taken from the literature review and help to find an approach to diversity measurement that is as holistic as possible:

R5. Three different dimensions of diversity are distinguished: variety, balance, and disparity. Each of these dimensions are necessary to describe a different aspect of diversity, which is why each of these dimensions alone is insufficient to measure diversity (Stirling, 2007). Therefore, all dimensions should be measured or carefully considered why they are not measured (Mcdonald & Dimmick, 2003; Stirling, 2007).

R6. In addition, Loecherbach et al., 2020 suggests measuring several contexts of diversity (entities, topics, viewpoints) to better understand how they interact. This is also necessary in terms of measuring content diversity as comprehensively as possible.

R7. According to Pariser, 2011, algorithmic suggestion should lead to a loop in which the user is only exposed to self-confirming content (Möller et al., 2018). Since this is the case in an individualized radio stream, the metric should be constructed in such a way that filter bubbles can be detected.

6.1.3 Characteristics of Radio

Furthermore, the following requirement is derived based on the specific characteristics of the radio from Hirschmeier et al., 2019, p. 5023. It must be

considered for the diversity measurement to maintain the originality of radio as a medium:

R8. “Mixture of formats like stories, articles, interview, long features and audio dramas” lead to an “diversified image”: The diversity metric should indicate if different formats are played out.

6.1.4 Individualized Radio

The last block of requirements results from the specifics of the individualized radio stream. These requirements must be considered to measure individualized radio streams in a practical, large-scale manner:

R9. The measure needs to be automatically applicable since many different personalized streams exist in parallel.

R10. Because many different individualized streams should be able to be represented together as a metric, every stream needs to be measured on its own (Janka, 2021) and it must be possible to interpolate the diversity metric. This results in several mathematical requirements:

- The metric should be monotonically increasing or decreasing for each of the three dimensions (variety, balance, disparity) (Laxton, 1978; Stirling, 2007). Theoretically, the diversity metric should increase/decrease linearly. Otherwise, when two metrics accumulate, one may have more influence on the overall value, even though they may both be very close. Therefore, the shape of the diversity curve (e.g., from 0 to 1) should be as uniform as possible (Stirling, 2007).
- Two equal subsets and objects or played contributions result in the same diversity value, independent of the played order (Laxton, 1978).
- In addition, a distribution of diversity metrics across users must be specified on the visual representation, as these are accumulated.

R11. Since the radio stream is infinite in theory, the measurement of diversity should always be up-to-date and not based on old streams.

R12. The metrics must be easy to understand so that the monitoring editors understand the metrics. Furthermore, it is important to be able to deduce from the metric how to act to adjust the current radio stream in the desired direction (Waldhauser, 2021).

6.2 Design Implications

This chapter makes suggestions on how diversity can be measured in individual radio, considering the requirements. Design implications are created for this purpose. All design proposals in this thesis assume that individual radio streams of users are measured. Thus, there is one stream or set of posts per user. These individual streams are to be measured and visualized in aggregated form. The design implications are split by implications that have a direct impact on the diversity measurement and implications that are relevant for displaying the information to the user and modifying or filtering the measurement for display.

6.2.1 Measurement Design

The following design suggestions relate to the diversity metric itself:

D1. Posts should be divided into different category systems:

i. Information, education, culture, and entertainment (R1).

ii. International, national, and regional (R2).

iii. Among others: Biography, economy, leisure, sports, religion, and environment (R2): Due to vagueness in the Public Broadcasting Act, it is not easy to interpret what is meant by "areas of life". But Deutschlandfunk already has a category system that can be described as "areas of life" which will be applied here (Janka, 2021).

iv. Left, mid left, center, mid right, right (R3): Looking at the political spectrum is one way to reflect the diversity of opinions and attitudes (Schulz et al., 2019). This can be measured both as an artificially created ordinal scale as in Donsbach. et al., 2015, or applied directly to the political spectrum right/left as in Schulz et al., 2019. In this paper, a midway approach will be

taken, dividing the political spectrum into five simple categories.

v. Among others: Interview, comedy, comment, news, and traffic (R8): Posts should be classified by format. Again, Deutschlandfunk already has an established system of categories for their radio formats. Deutschlandfunk has been a classic radio station for many years (Deutschlandradio, 2021c) and therefore has expertise in this area on which formats are relevant for radio. Their classification scheme is applied here.

vi. Politicians, government, companies, journalists, experts, ordinary citizen, NGO, religion, reformist group, judicial, police/executive, anonymous: The contexts of diversity, viewpoint (diversity of opinions) as well as topic (diversity by "areas of life"), are already covered by the above classifications. Requirement R6 proposes to measure all contexts of diversity. Thus, a further classification according to the dimension entity is considered. This indirectly fulfills R3 as well since a diversity of sources is also related to different opinions and viewpoints (Bagashka, 2014). As there is no classification scheme available at Deutschlandfunk, I derived the classification from various works on it: The above entities are types of sources that are used similarly in most works. Swert & Hooghe, 2010 and Armstrong, 2006 classify by gender. Freedman et al., 2007 use the actual named entities to classify contributions. Bagashka, 2014; Benson, 2009; Masini & van Aelst, 2017; Matthews, 2013; Saridou et al., 2017; Van Leuven et al., 2015; Yoon, 2013; Young & Dugas, 2012, on the other hand, use similar classifications to the social groups listed above. Moreover, in some cases, they additionally use specific groups that fit the subject they are studying, such as "immigrant" (Benson, 2009).

According to R5, all dimensions of diversity should be measured. In this case, the explicit measurement of the dimension disparity is omitted. But the choice of categories already significantly influences

disparity (Stirling, 2007). May, 1990 and Stirling, 2007 even argue that variety and balance cannot be measured without first characterizing disparity. Category systems are created on the basis of disparity (May, 1990; Stirling, 2007). The diversity of classifications and the choice of categories in this case already represent a very broad spectrum. Especially for the categorization of the political spectrum, whose categories can be represented on a scale. The measurement of disparity for the political spectrum would have been particularly obvious. But the added value of this additional diversity metric, which plays against simple understanding (R12), is in question since the categories are already well distributed.

D2. Measure the dimension balance: For requirement R1 to R3 it would be sufficient to measure the number of categories (variety). But by requirement R4, the dimension balance is mentioned additionally. Therefore, the measurement must cover both dimensions.

D3. The category systems should be measured with the relative entropy Index (R1 - R4): Measurement variables that already reflect both dimensions in one key figure are particularly suitable for this purpose. This reduces complexity by omitting a further key figure, which in turn satisfies R12. The less overloaded the visual representation is, the easier it is to understand.

As open diversity and ordinal scale cannot measure both dimensions at the same time they are omitted as metrics. Reflective diversity (as a metric) and the Deviation index are also dropped because they do not fit the deliberative normative framework and use target values for calculation. In the remaining selection of metrics, the Simpson index and the Shannon entropy remain in their various forms.

The Shannon entropy as well as the Simpson index can both be considered for diversity measurement in individualized radio. Because both have a normalized form, both are monotonically increasing/decreasing, and both are resulting in the same diversity value for equal sets of elements (R10). Furthermore, both are quantitative

metrics and, assuming that the metadata for the radio posts are available for all categorizations, they are automatically applicable (R9).

In the optimal state, the diversity curve would have a linear shape (Boydstun et al., 2014) to satisfy the R10 requirement of being interpolable particularly well. Shannon entropy and Simpson index are both non-linear curves. But in contrast to the Simpson index, the Shannon entropy is overall more sensitive to changes. Especially in the maximum and minimum value ranges, the Simpson index is rather inelastic due to its use as a market concentration index. The Shannon entropy is also better in this range (Boydstun et al., 2014). Therefore, the Simpson index is dropped from the selection.

To satisfy requirement R10 with the Shannon entropy, a uniform scale of values is required. This is necessary because different category systems with different numbers of categories emerge from the requirements. It also creates clarity and a simpler understanding when key figures from different category systems are compared with each other (R12). The standard variant as well as the entropy in number equivalents cannot be applied in this case. Therefore, relative entropy is chosen as the metric for diversity measurement in this work.

D4. The individual diversity measures should not be scaled. That means, a stream from user A should be included in the overall metric with the same weight as a stream from user B. Even if user A has listened to twice as many posts or minutes. This will serve to ensure equality among users. For diversity to be played out to society, every user counts equally.

Also, when calculating at the user level, the length of posts in minutes should have no effect on user's diversity. Only the number of posts per category should have an effect. Because it is not necessarily the length that counts, but whether a post of these categories is played at all (Bittner, 2021).

Filter bubbles are avoided (R7): By using many different categorization systems and thus already covering all dimensions (at least indirectly) and all diversity contexts (R7).

6.2.2 Display Design

The following design suggestions relate to the visualization as well as the associated filtering of the underlying diversity metric:

D5. Filter the streams heard on a timeline (R11): To set filter on current streams.

D6. Option to display the distribution of diversity across users (R10): since the diversity value is displayed aggregated for all streams, there should also be an overview over the diversity distribution.

D7. Display distribution of category systems (R12): From the actual diversity metric it is not possible to see where the distribution deviates. To draw recommendations for action from the diversity display, the distribution of the category systems should be displayed.

D8. A distinction should be made between diversity played out and diversity heard (Janka, 2021). Deutschlandfunk can thus distinguish whether they are fulfilling their mandate (from the Public Broadcasting Act) by the played-out diversity. In contrast, they can check whether the user receives his/her diversity differently by skipping posts.

D9. Diversity that the user has heard: Posts should not be included in the diversity metric until a certain amount of listening time has been reached (Barknecht, 2021; Bittner, 2021).

6.3 Dashboard Mockup Proposal

A dashboard is used to visualize the diversity indicators: "These tools help people visually identify trends, patterns and anomalies, reason about what they see and help guide them toward effective decisions" (Brath & Peters, 2004, p. 1). It therefore supports R12 and is chosen as the visualization option for this work.

A dashboard is used to present information in an aggregated visual format (Few, op. 2006). The information can be used for decision-making as well as for individual goals in operational work (Few, op. 2006). Dashboards have evolved from

static one-page views to multi-page views with interactive elements (Few, op. 2006; Sarikaya et al., 2018). However, there are still both types, which can be divided into functional and visual, with functional being the new type of dashboard that is interactive (Sarikaya et al., 2018).

The design of a dashboard is influenced by its users (Sarikaya et al., 2018). Two different user groups with different knowledge levels must be considered for Deutschlandfunk. On the one hand, editors who monitor the individualized radio streams for their diversity, and on the other hand, the listeners, who are to be shown reduced diversity information for the traceability of their personal radio stream. For this reason, two different dashboards are created.

6.3.1 Editors

The editors deal with the diversity dashboard daily. They also need to be able to filter by time and other factors to monitor the individualized streams. Therefore, a functional dashboard is created for the editors with various filtering and analysis features. The editors' workstation is the computer, so the dashboard is mocked up in computer screen size. The following figure shows the standard view of the dashboard. Other views are shown in appendix B Dashboard Mockups.

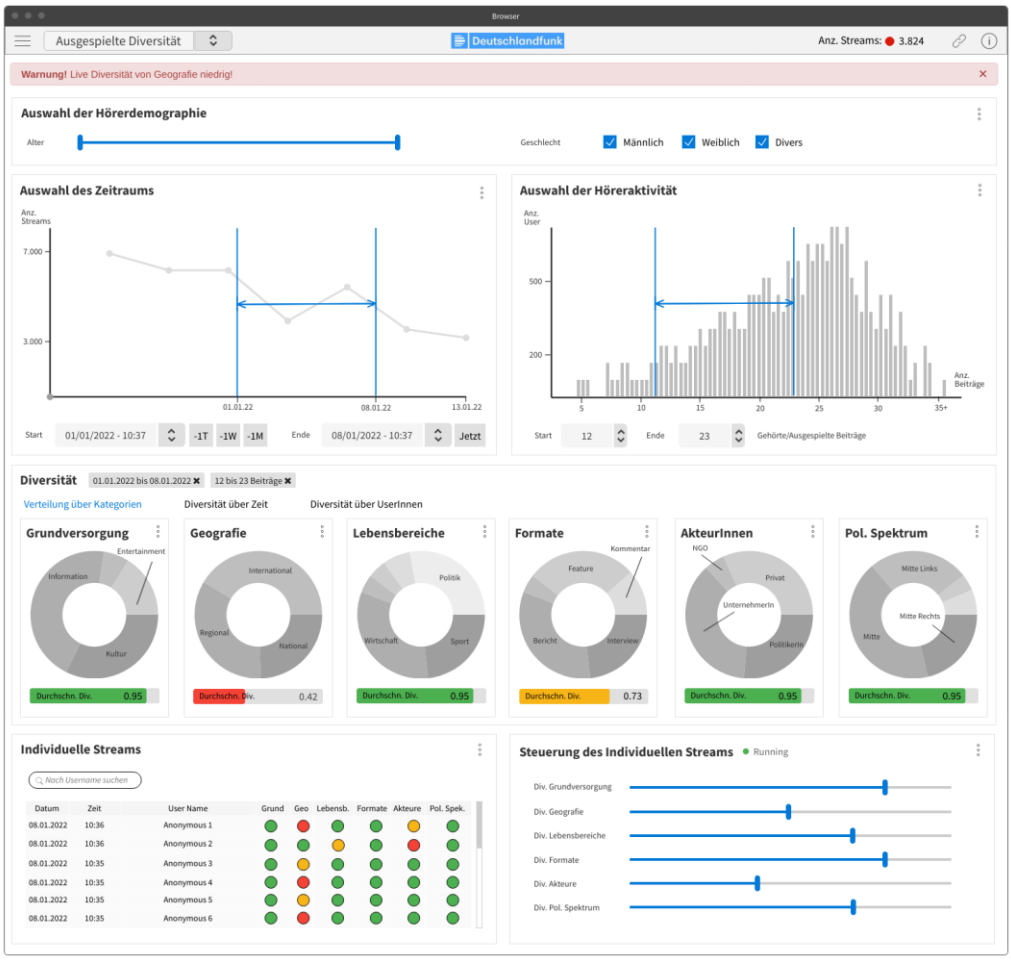

Figure 4: Editors Dashboard - Category Distribution View

On the top menu bar, the user can set whether he wants to display the diversity played out ("Ausgespielte Diversität") or the diversity heard by the listener. Below that are the setting options for listener demographics ("Hörerdemographie"), time period ("Zeitraum") and listener activity ("Höreraktivität").

In the listener demographics section, the streams can be filtered for age and gender of the listeners. In the time period section, the streams can be narrowed down in time. To the right of the time period section is the listener activity section. Here, the user can set the minimum number of listened elements a stream should be included in the diversity calculation. This prevents very short streams, for which the diversity calculation does not make sense, from being included in the calculation (D9).

Below the filter options is the area that displays the diversity ("Diversität"). For each category system, a diagram is displayed that graphically shows the distribution of the respective category. The diversity (relative Shannon entropy) is displayed below as a value from 0 to 1. As a colored indicator, the diversity value is marked

green/yellow/red on a progress bar to make it immediately apparent when the value falls into critical ranges.

Furthermore, it is possible to display different diagrams for each category system within 3 sub-menus in the diversity section. These are the buttons "Distribution over Categories" (“Verteilung über Kategorien“), "Diversity over Time" („Diversität über Zeit“) and "Diversity over User" (“Diversität über UserInnen”). "Distribution over Categories" is already described above. "Diversity over Time" shows the diversity value from 0 to 1 over a period of time. "Diversity over User" shows how diversity is distributed over the number of users. Both can be seen in the appendix B Dashboard Mockups.

At the bottom left of the dashboard is another filtering option. Individual streams can be manually selected to examine diversity. To the right is an indication of how a possible integration of the recommender system into the dashboard might look. For example, as here, one slider per category system can be controlled.

6.3.2 Listeners

A very important playback medium of the personalized radio is the smartphone, which also provokes the greatest possible compactness of the diversity display. Furthermore, the listeners should be given a simple overview of the composition of their personalized stream. For this reason, a visual dashboard is chosen for the listener dashboard without the requirement of domain knowledge.

Two options were modeled and are shown in Figure 5: Listeners Dashboard Options. On the left side, sentences that allow for some fuzziness in the representation. A school grade is assigned per category system and indicates how diverse the listener hears. It is not specified at which diversity value these limits of school grades take effect. On the right side, a display of the distributions per category system. In neither option the user is shown the diversity score.

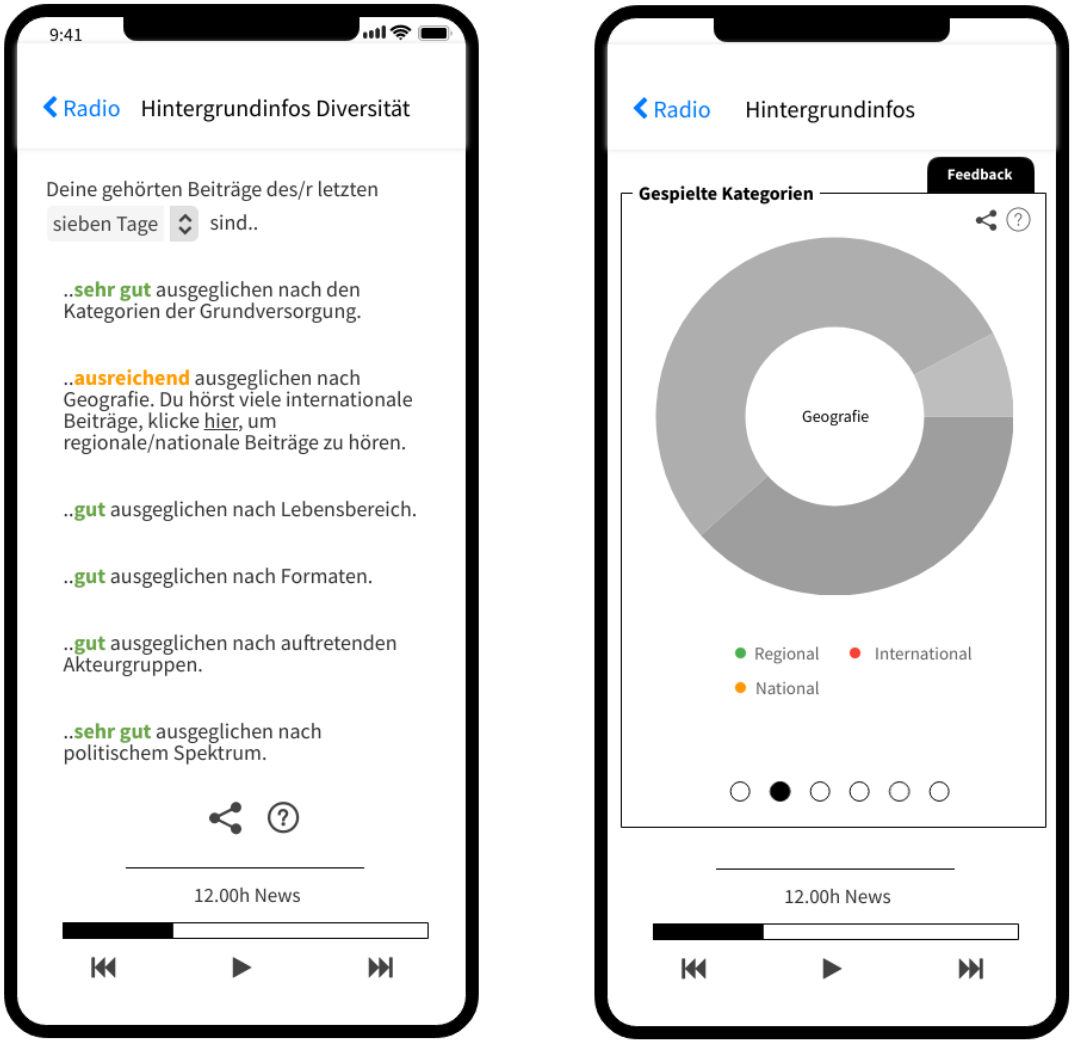

Figure 5: Listeners Dashboard Options

6.4 Evaluation with Deutschlandfunk

Four informal conversational interviews (Turner III, 2010) are conducted with Deutschlandfunk. The interviews are conducted with different roles (online editorial, IT, program management). In each interview, the dashboards are presented, and a brief explanation of the diversity metric is given. The interviews are structured in such a way that the interview partners can give feedback at any time. The direct structure of the interview should allow participants to provide "in the moment experiences" and feedback (Turner III, 2010). During the interviews, questions arose for which user group the browser dashboard can be used and what actions can be derived from the information gained. These questions were further pursued in the respective next interviews. The feedback of Deutschlandfunk was used to manifest and add to the design implications as well as the dashboard. It is summarized below:

Remark	**Implemented**
(Janka, 2021): Swap axes for diagram user distribution (diversity was on y axis).	Yes
(Janka, 2021): Distinguish between played out diversity and diversity that the user experiences by skipping posts in the individual stream.	Yes
(Bittner, 2021): In addition to played out diversity and user heard diversity, another option: diversity produced by all editors. This can be used to check whether the appropriate diversity played out in individualized radio is present in the produced content.	No, as the dashboard monitors the personalized radio. Future adjustments with another view and more filters are possible.
(Janka, 2021): Diversity of streams shall be measured individually. Only after that, the diversity metric should be accumulated. This prevents balancing of contribution categories of users among themselves.	Yes
(Bittner, 2021; Janka, 2021): More intuitive captions/dashboard.	Yes, but less than requested. Other descriptions and simplifications were weighed against clarity and functionality.
(Janka, 2021): In the distribution graphs tab, replace pie charts with bar charts. Advantage: Proportions can be put into relation faster. Disadvantage: GUI becomes more confusing.	No, would have made the dashboard cluttered.
(Janka, 2021): Do not apply filters in the distribution graphs tab to see distributions across all users.	No, the distribution tab should show the distribution used as the basis for the

Remark	Implemented
	diversity metric. There could otherwise be confusion if the diversity metric and distribution show different diversities.
Call-to-Action and who should use it: (Janka, 2021): Recommender system control should not be displayed in editorial. Editors should only get the dashboard to monitor diversity. On a regular basis, for example once a week, the IT could coordinate with the editorial teams and intervene if action is needed, such as adjusting the algorithm or its parameters. (Waldhauser, 2021): Editorial offices are decentralized, the dashboard does not fit there, because editorial offices have the focus on their program. Perhaps a central office would have to be created to monitor the dashboard. (Barknecht, 2021): Typical users would be online editors who take apart the linear program and reassemble it (have good overview), data analysts and IT to control the recommender system, "audience development" who investigate specific questions about the current program, and the intendant to control whether the guidelines of the Public Broadcasting Act are being followed. The dashboard would be used more passively as a source of reporting. (Bittner, 2021): Recognizes different user groups and proposes to create a simple and an advanced dashboard. This allows editors with individual shows to check their assumptions about the diversity of their content.	The dashboard was developed with the purpose of making it available to editors who have an overview of all programs. Nevertheless, it is component-based and can be customized. It can be used unchanged for all central offices for analysis/display. Possibly the control of the recommender system should be switched off. Should it be used in individual editorial offices that create certain programs, it can be tailored to their needs.

Remark	**Implemented**
(Barknecht, 2021; Waldhauser, 2021): It is tricky to adjust diversity for individual users, as the user may skip some posts for good reasons.	No, as the focus of this work is on diversity measurement. Needs to be discussed when implementing/integrating the recommender system with the diversity dashboard.
(Bittner, 2021): Anonymize usernames for data protection.	Yes, it is irrelevant for the diversity calculation to display the usernames.
(Waldhauser, 2021): Equal distribution vs. station-specific distribution: the distribution of categories should not always be equal but is coordinated within the station. This must also be adjustable. (Waldhauser, 2021): Questioning the role of the media or reference to normative frameworks: Is it the media's job to present even small fringe opinions? And if so, how?	No, target values mean reflective diversity, which is beyond the scope of this work and addresses a different normative framework (Liberal Aggregative). This could be part of a future work.
At what point is a contribution counted as heard?: (Barknecht, 2021): A contribution is considered to have been heard from around one minute onwards. There should be a slider where you can set the length of a post to be counted as heard. In seconds as well as in heard percentages of the article. (Bittner, 2021): Fixed value 60 seconds or more of listening time should classify a post as heard (international standard).	Yes, a minimum listening time for a post is considered, however one should be able to manually adjust this in the background to seconds as well as percent of post.

Remark	Implemented
(Bittner, 2021): There needs to be a minimum listening time from which diversity can be measured at all. Few listened content cannot be diverse.	Yes, a filter on minimum/maximum posts listened is already implemented.
(Barknecht, 2021): Contrast between diversity and number of listeners. If the diversity is adjusted on the sliders for a user, it must also be tracked how the listener behavior of the user changes.	No, as the focus of this work is on diversity measurement. It is important in creating a dashboard that embeds a recommender system and allows the control of individual streams.
(Bittner, 2021): There should be the ability to filter by listener demographics such as age and gender to examine diversity by listener group.	Yes, dividing by listener groups supports diversity measurement as it is possible to examine additional levels

Table 2: Feedback Interview Browser Dashboard

Remark	Implemented
Sentences vs. charts option: (Janka, 2021): Qualitative data is better, more can be read from diagrams than from sentences. (Waldhauser, 2021): Rather soft phrases. Sentences. (Barknecht,	Both options are given.

2021): Skepticism towards pie charts in mobile view, as they pretend exactness. The classification of posts is not always clear. Therefore, rather sentences that consider diversity.	
(Waldhauser, 2021): Probably not used by users. Previous user tests have shown that few users click on further information. (Barknecht, 2021): No concerns about text not being read.	-
(Barknecht, 2021): Possibly the information of diversity unsettles the user - also with regard to the fact that the user realizes that the stream is compiled by an algorithm. This should be investigated in a user study beforehand.	No, a user study can be conducted before implementation.
(Barknecht, 2021): Dashboard should include a link to Public Broadcasting Act information on why the information is being displayed.	Yes, link button with further information is already provided in the mockup.

Table 3: Feedback Interview Mobile Dashboards

6.5 Application on Test Data

Deutschlandfunk has provided a dataset of played elements from regular radio (non-personalized radio). This describes all played elements of two days for each of the three stations (Deutschlandfunk, Deutschlandfunk Nova and Deutschlandfunk Kultur). One weekday (02.11.2020) as well as one Sunday (31.10.2020) are included. For each played element, the format, the geographical information and the "area of life" are provided. A post can have multiple categories from a category system. For example, a post can contain politics and economics. These are both considered in the following calculation. The post would therefore contribute to the category politics and economy. Since it follows from design implication D4 that a contribution counts as heard if it was simply played out, it can be assumed that a post that contributes to different categories also contributes to all categories in the diversity calculation.

Relative entropy was calculated for each of the classifications available for this data set (1 represents a perfect distribution, 0 the worst distribution):

Date/Station	**Deutschland-funk**	**Deutschland-funk Nova**	**Deutschland-funk Kultur**	**Total**
31.10.2020	0,8104	0,8431	0,7740	0,8440
02.11.2020	0,7909	0,8027	0,7150	0,8118
Total	0,8185	0,8511	0,7593	0,8385

Table 4: "Area of life" - Relative Entropy

Date/Station	**Deutschland-funk**	**Deutschland-funk Nova**	**Deutschland-funk Kultur**	**Total**
31.10.2020	0,9694	0,9792	0,9856	0,9800
02.11.2020	0,9389	0,9634	0,9272	0,9428
Total	0,9531	0,9723	0,9662	0,9628

Table 5: Geography - Relative Entropy

Date/Station	Deutschland-funk	Deutschland-funk Nova	Deutschland-funk Kultur	Total
31.10.2020	0,5047	0,4230	0,5462	0,5432
02.11.2020	0,4822	0,4095	0,5156	0,5087
Total	0,5042	0,4315	0,5404	0,5335

Table 6: Format - Relative Entropy

The total values are not averages, but the relative entropy for the entire set in each case. This can be observed in the example of the category system "area of life" of Deutschlandfunk Nova: The overall diversity is higher (0.8511) than the diversity of the individual days (0.8431 and 0.8027). The two sets balance each other and cumulatively contribute to greater diversity. There may be a bias due to unequally distributed sets, but the time period is the same for all stations, the sets should therefore be roughly balanced in size.

It is apparent that the diversity for each station per category system is similar. For example, all stations have a diversity of >0.95 for geography. The diversity for the weekend day (31.10.2020) is higher than for the weekday in all category systems. Thus one can already recognize patterns in the different category systems even in this small sample. They can be identified and could now be followed up in a real scenario.

6.6 Concept for the Implementation at Deutschlandfunk

This chapter briefly describes the steps Deutschlandfunk needs to take to introduce the Diversity Dashboard. The information comes from the conversations with Deutschlandfunk.

1) As a first step, Deutschlandfunk must collect the necessary metadata for the contributions. These are already available for four of the six category systems. It is also possible to initially create the dashboard with a smaller number of category systems.

2) The second step involves selecting the user groups/entities to whom the browser dashboard will be made available. It must be clarified which user group should use which components. In addition, it should be defined for which purpose, and with which recommended action the dashboard is to be used in each case.

3) The third step is to implement the browser dashboard and make it available to the users.

4) After implementation, it should be checked how the user groups interact with the dashboard and how it is perceived.

5) As a final step, the listener dashboard can be implemented in the radio app. Before implementation, Deutschlandfunk must decide on one of the two visualization options.

7 Discussion and Limitations

In this work, diversity metrics from the media were gathered using a literature review and applied to personalized radio. Using requirements from the Public Broadcasting Act, literature, and radio specifics, the metric was chosen and a dashboard to monitor diversity was developed. The relative Shannon entropy was selected as the diversity metric. It is particularly suitable for measuring diversity because it has a uniform scale. It is also easy to understand and can be automated for individual streams. Compared to the Simpson index, which is also suitable, it has the advantage that it is more elastic and therefore reacts better to changes in diversity. Furthermore, Shannon entropy is one of the most widely used metrics in the media for content diversity, which is why it is also particularly useful for an initial diversity measurement in the context of personalized radio. Unlike other metrics, Shannon entropy can combine the dimensions of variety and balance into one metric. To comprehensively measure diversity, six different category systems (e.g. radio format, political spectrum or “area of life”) were used. For each of which the relative Shannon entropy is measured and displayed on the dashboard.

Feedback from experts of Deutschlandfunk was incorporated into the diversity metric and the dashboard. For example, that streams should be measured individually per user before being interpolated. The feedback that was not implemented is mostly beyond the scope of this work but is extremely helpful for future research towards implementation or extension of the dashboard. In addition, the interviews with Deutschlandfunk started many discussions like the possible uses of the dashboard or the actual definition of diversity.

Measuring diversity in personalized radio is a novel approach. So far, diversity has only been measured in linear radio. The feedback from the experts at Deutschlandfunk suggests that the dashboard provides a good way to quantify the diversity of individualized radio. It can be used to monitor compliance with the Public

Broadcasting Act and report diversity. Based on many issues that have been raised in the interviews, this will lead to further discussions and research opportunities. In addition, by gathering diversity types, diversity metrics, requirements for measurement, and design implications for implementing a diversity metric and visualization for personalized radio, this work provides an overview of what is available. Other researchers and radio stations can use this as a foundation, as well as decide for themselves whether to modify the metric and the dashboard. Future research can address the implementation of the dashboard.

The following paragraphs discuss the limitations and to what extent the results can be generalized.

In general, it must be considered that determining how diversity should be measured is a certain kind of power (Karppinen, 2007). The one who sets the standards indirectly determines what should be played on the radio. This is always subjective (Stirling, 2007). Since Germany is a democracy and there is a consensus to support freedom of expression, the German states have passed the Public Broadcasting Act. In this work, most of the requirements come from the Public Broadcasting Act. It is intended to promote free speech and a democratic system. Therefore, the diversity metric only holds its validity in this democratic system. In societies with other values, other requirements may need to be considered.

Juraitè, 2014 argue that diversity cannot be measured quantitatively because of social and cultural complexity. This is important to consider for the dashboard of this thesis. A diversity measurement cannot be complete. In this work there are different categorizations for radio contributions. Derived from the Public Broadcasting Act, from literature, from radio specifics or from already existing schemes at Deutschlandfunk. It must be noted that there are numerous other classifications of contributions that have not been applied. Thus, this diversity metric is not complete. It cannot measure all dimensions and all contexts in all their categorizations. However, it can cover them as well as possible. This should be assumed for the dashboard created. With incompleteness comes the fact that it is probably not possible to avoid all filter bubbles and echo chambers. However, the most important categorizations of the Public Broadcasting Act are covered. For example, the modeled political spectrum already prevents the socially popular division into left and right. When adjusting the dashboard, it should be checked or made aware at each point in time, which dimension

and which context of diversity is changed as the dashboard is designed to represent most dimensions and contexts of diversity.

Regarding Deutschlandfunk, it might be better to leave out the small editorial teams that produce individual programs based on the diversity measurement and let them evaluate their work at a more qualitative level. The central office that monitors the dashboard could then only intervene in an emergency after a qualitative analysis. Thus, there is no danger of individual editorial teams limiting themselves to the diversity of the dashboard. On the other hand, the dashboard could serve as a good validation opportunity for small editorials to check their own assumptions. This could lead to higher diversity in other dimensions.

The currently modeled dashboard gives a way to monitor actual diversity. During the interviews, it also became clear that it must also be possible to specify target values. These should reflect the proportions of a category system set by Deutschlandfunk. For example, it could be set up that the target share of entertainment is 30%, whereas culture should only take up 20%. This is the concept of reflective diversity. Furthermore, under the condition that the editors already create a diverse offer, it would be possible to assess how the personalized stream compares to the created posts by the editors in terms of diversity, too. Here, future research could implement the dashboard for reflective diversity. But the Public Broadcasting Act only mentions the balance of posts as a requirement, so it is not clear how diversity is to be measured or how the target percentages can be changed. Any metric created, as well as the dashboard of this paper, will need to be reviewed by legal professionals to determine if and to what extent the dashboard indicates compliance with the Public Broadcasting Act. From this perspective, this dashboard is a good starting point to spark the discussions and legal issues. Future research can address how reflective diversity can be integrated into measurement and what target values may be used to measure it.

From the four normative frameworks, the deliberative framework was chosen to capture actual diversity values by categories with equal shares. With advancing technology, for example, the introduction of recommender systems or of diversity measurements, there is an attempt to consider the constructs used there, such as the normative framework, as superior (Karppinen, 2006). Therefore, it is important to state that other normative frameworks can be considered in the context of diversity, too. These, of course, serve a different purpose than the assessment of the individual radio,

which is done within the deliberative framework in this thesis. The recording of target values as in above paragraph about reflective diversity or the highlighting of minorities in the adversarial framework may be just as important in a democracy and may also be necessary to solve social conflicts. Further research would therefore to be sought for the other frameworks.

What is most important when measuring diversity in practice is the existence of metadata. That is because diversity cannot be measured without the individual radio posts being linked to their categories. For example, it might not be possible to map the political spectrum in the metadata of the posts, and a broadcaster might decide to use a different proxy for diversity of opinion. The basic functions on the dashboard such as diversity measurement and the display of the different category systems with distribution would remain the same in that case, only individual components could be exchanged or changed.

The evaluation with Deutschlandfunk carries the risk of bias, since only one station group is considered. The created dashboard including the selected metric, however, is mostly derived from the Public Broadcasting Act and only a small part from the interviews with Deutschlandfunk. Therefore, it can be assumed that the dashboard can be used for other radio stations as well. Particularly noticeable is that there are various user groups for the dashboard at Deutschlandfunk and that there are editorial teams that only create one program without an overview over all the other programs. For this, it is also possible to customize the dashboard and omit individual elements. In this way, it could also be adapted for the requirements of other broadcasters.

As described above, it turned out during the interviews that there are not "the editors" at all, but many different editorial teams that are responsible for single or multiple broadcasts (Waldhauser, 2021). Therefore, a central position needs to be created to oversee the dashboard for an overall diversity. This can be both the intendant to check compliance with the Public Broadcasting Act or the IT, which for example can change the algorithm of the Recommender System as a direct "call-to-action". For the smaller decentralized editorial teams that create individual broadcasts, the dashboard can be customized and used to check their own assumptions. For example, an editorial department could check the extent to which it reports in a balanced way according to political spectrum only.

Performance must be considered when implementing the dashboard in practice. This is because, due to its interactivity, for example a time filter can be applied to the stream. The entropy of all streams in the filter would have to be recalculated. The performance impact has not been studied and would need to be considered in the implementation.

The research methodology used has resulted in a valid diversity metric that can be used to measure diversity in individualized radio as well as at Deutschlandfunk. The literature review with concept matrix has been the right tool to collect and classify the different metrics as it created a good overview and allowed to classify the approaches during the research. For collecting the requirements for the diversity metrics, it would have been better to involve the interview partners already during the collection of the requirements, as this would have meant that no retroactive adjustments would have had to be made and there would have been earlier feedback. The evaluation with the interview partners did not begin until after the dashboard had been created. Nevertheless, most of the assumptions that were made were also confirmed in the interviews with Deutschlandfunk. Particularly beneficial to this work was the mixed method approach, which allowed the diversity metric to be taken from theory directly into a visualization that can be used in practice. This has also triggered further discussions at Deutschlandfunk and has shown boundaries like non-existing metadata of the diversity metric when applied in a practical context. These could be incorporated directly back into the design of the dashboard.

8 Conclusion

Diversity is essential to democratic discourse and is recognized as such by the Public Broadcasting Act. Public broadcasters must play out a certain diversity. This is currently not measured in personalized radio. This work therefore aimed to answer the research question "How can content diversity be measured and quantified for personalized radio broadcasts?" and therefore developed a framework to measure content diversity in individualized radio. The framework suggests a way to measure diversity in individualized radio.

To do this, a literature review focusing on works that measure diversity in the media is conducted and applied to personalized radio. Particularly frequently used metrics are the Simpson index and Shannon entropy. These measure how evenly

distributed posts are across categories in a given category system (e.g., politics, religion, environment). They also take into account the number of categories given. In the literature, the metrics are also typically scaled by the length of the posts.

Following the literature review, the requirements for diversity measurement are established and the appropriate metric is derived based on these. The requirements are largely derived from the Public Broadcasting Act, the literature, and characteristics that must be considered in individualized radio. Specifically, this thesis concludes that diversity should be measured in six different category systems using the relative Shannon entropy:

- Basic supply (information, education, culture, entertainment)
- Geography (international, national, regional)
- Areas of life (biography, economy, leisure, sports, etc.)
- Formats (interview, comedy, comment, news, traffic etc.)
- Actors (politicians, government, companies, journalists, etc.)
- Political Spectrum (left, mid left, center, mid right, right)

Based on this, a dashboard mockup is created that visualizes diversity in the six category systems. It also includes filtering options to narrow down diversity by certain characteristics such as time, individual users, and listener demographic information. The dashboard is modeled as a browser version for editors and a simplified version for listeners as a mobile version. The dashboards and diversity metrics are further evaluated with interview partners from Deutschlandfunk and adapted according to the feedback. Among other things, a filter for listener demographic information was added. It is also suggested that the components of the dashboard should be adapted to the respective user group at the broadcaster or Deutschlandfunk, since individual editorial teams would only use parts of the dashboard.

When applying relative entropy to test data from a normal, non-individualized, radio stream, patterns could already be detected. For example, diversity was higher on weekend days than during the week.

The transformation from linear radio to individualized radio is accompanied by many challenges. Public broadcasters must pay particular attention to continuing to comply with the Public Broadcasting Act and its resulting requirements. This work

contributes to the understanding of the whole context of individualized radio and diversity. It also offers a concrete proposal to measure diversity in personalized radio. The dashboard provides an opportunity to quantitatively assess the requirements of the Public Broadcasting Act.

For radio broadcasters, this work helps them address the issue of diversity and question important issues such as the underlying normative framework. Furthermore, they can implement the dashboard mockup to measure the diversity of their individual streams.

The created dashboard as well as the relative Shannon entropy are good indicators to determine the current state of diversity and for example report on it. But a practical test is still needed, to find out to what extent the diversity metric can report on diversity and justify the played-out streams. Furthermore, it is not certain whether all the data required for the diversity metric can be collected in a practical application.

The requirements that were established and the interviews that were conducted are specific to the Public Broadcasting Act and possibly to the context of Deutschlandfunk. These are based on the goal of the deliberative normative framework, which also records only one of four normative perspectives.

Future research should implement the dashboard and examine the extent to which the Public Broadcasting Act can be justified in a practical application. In addition, further research on the integration of target values and the associated implementation of reflective diversity in the dashboard would be desirable. As part of the implementation, it can also be investigated whether and how a recommender system can be meaningfully integrated into the dashboard so that it is possible to control the recommender system from the dashboard.

The integration of a dashboard for the listeners of the radio is a possibility for further research, especially regarding the acceptance and effect of such a dashboard.

References

Agresti, A., & Agresti, B. F. (1978). Statistical analysis of qualitative variation. *Sociological Methodology*, *9*, 204. https://doi.org/10.2307/270810

Ahern, S. (2011). *Making Radio: A Practical Guide to Working in Radio in the Digital Age* (3rd ed.). Allen & Unwin. https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=5434526

Alexander, P. J., & Cunningham, B. M. (2004). Diversity in broadcast television: An empirical study of local news. *International Journal on Media Management*, *6*(3-4), 176–183. https://doi.org/10.1080/14241277.2004.9669400

Allen, H., Connolly, S., & Hargreaves Heap, S. P. (2017). Media pluralism: What matters for governance and regulation? *Journal of Media Economics*, *30*(2), 47–64. https://doi.org/10.1080/08997764.2017.1364257

An, J., Quercia, D., Cha, M., Gummadi, K., & Crowcroft, J. (2014). Sharing political news: The balancing act of intimacy and socialization in selective exposure. *EPJ Data Science*, *3*(1). https://doi.org/10.1140/epjds/s13688-014-0012-2

Armstrong, C. L. (2006). Story genre influences whether women are sources. *Newspaper Research Journal*, *27*(3), 66–81. https://doi.org/10.1177/073953290602700306

Aslama, M., Hellman, H., & Sauri, T. (2004). Does market-entry regulation matter? *Gazette (Leiden, Netherlands)*, *66*(2), 113–132. https://doi.org/10.1177/0016549204041473

Avraham, E., & First, A. (2010). Can a regulator change representation of minority groups and fair reflection of cultural diversity in national media programs? Lessons from israel. *Journal of Broadcasting & Electronic Media*, *54*(1), 136–148. https://doi.org/10.1080/08838150903550469

Baden, C., & Springer, N. (2014). Com(ple)menting the news on the financial crisis: The contribution of news users' commentary to the diversity of viewpoints in the public debate. *European Journal of Communication*, *29*(5), 529–548. https://doi.org/10.1177/0267323114538724

Bae, H.-S. (2000). Product differentiation in national tv newscasts: A comparison of the cable all-news networks and the broadcast networks. *Journal of Broadcasting*

& Electronic Media, *44*(1), 62–77. https://doi.org/10.1207/s15506878jobem4401_5

Bagashka, T. (2014). The bulgarian “nuclear referendum” of 2013 and the independence of the bulgarian media. *East European Politics*, *30*(4), 568–587. https://doi.org/10.1080/21599165.2014.960520

Barknecht, F. (2021, July 2). Interview by F. Terhalle, & S. Hirschmeier.

Baum, M. A. (2013). The iraq coalition of the willing and (politically) able : Party systems, the press, and public influence on foreign policy. *American Journal of Political Science*, *57*(2), 442–458. https://doi.org/10.1111/j.1540-5907.2012.00627.x

Bechmann, A., & Nielbo, K. L. (2018). Are we exposed to the same “news” in the news feed? *Digital Journalism*, *6*(8), 990–1002. https://doi.org/10.1080/21670811.2018.1510741

Benson, R. (2009). What makes news more multiperspectival? A field analysis. *Poetics*, *37*(5-6), 402–418. https://doi.org/10.1016/j.poetic.2009.09.002

Benson, R., Neff, T., & Hessérus, M. (2018). Media ownership and public service news: How strong are institutional logics? *The International Journal of Press/Politics*, *23*(3), 275–298. https://doi.org/10.1177/1940161218782740

Berry, S. T., & Waldfogel, J. (2001). Do mergers increase product variety? Evidence from radio broadcasting. *The Quarterly Journal of Economics*, *116*(3), 1009–1025. https://doi.org/10.1162/00335530152466296

Binderkrantz, A. S [Anne Skorkjaer] (2012). Interest groups in the media: Bias and diversity over time. *European Journal of Political Research*, *51*(1), 117–139. https://doi.org/10.1111/j.1475-6765.2011.01997.x

Binderkrantz, A. S [Anne Skorkjær], Bonafont, L. C., & Halpin, D. R. (2017). Diversity in the news? A study of interest groups in the media in the UK, Spain and Denmark. *British Journal of Political Science*, *47*(2), 313–328. https://doi.org/10.1017/S0007123415000599

Bittner, B. (2021, July 6). Interview by F. Terhalle, & S. Hirschmeier.

Borum Chattoo, C., Aufderheide, P., Merrill, K., & Oyebolu, M. (2018). Diversity on U.S. Public and commercial tv in authorial and executive-produced social-issue

documentaries. *Journal of Broadcasting & Electronic Media*, *62*(3), 495–513. https://doi.org/10.1080/08838151.2018.1451865

Boydstun, A. E., Bevan, S., & Thomas, H. F. (2014). The importance of attention diversity and how to measure it. *Policy Studies Journal*, *42*(2), 173–196. https://doi.org/10.1111/psj.12055

Bozdag, E., Gao, Q., Houben, G.-J., & Warnier, M. (2014). Does offline political segregation affect the filter bubble? An empirical analysis of information diversity for dutch and turkish twitter users. *Computers in Human Behavior*, *41*, 405–415. https://doi.org/10.1016/j.chb.2014.05.028

Bozdag, E., & van den Hoven, J. (2015). Breaking the filter bubble: Democracy and design. *Ethics and Information Technology*, *17*(4), 249–265. https://doi.org/10.1007/s10676-015-9380-y

Brath, P., & Peters, M. (2004). *Dashboard design: Why design is important*. http://cs.furman.edu/~pbatchelor/csc105/articles/TUN_DM_ONLINE.pdf

Bundeszentrale für politische Bildung. (2016). *Duden Wirtschaft von A bis Z: Grundlagenwissen für Schule und Studium, Beruf und Alltag* (6th ed.).

Carlson, M. (2007). Order versus access: News search engines and the challenge to traditional journalistic roles. *Media, Culture & Society*, *29*(6), 1014–1030. https://doi.org/10.1177/0163443707084346

Carpenter, S. (2010). A study of content diversity in online citizen journalism and online newspaper articles. *New Media & Society*, *12*(7), 1064–1084. https://doi.org/10.1177/1461444809348772

Chambers, T. (2003). Radio programming diversity in the era of consolidation. *Journal of Radio Studies*, *10*(1), 33–45. https://doi.org/10.1207/s15506843jrs1001_5

Champion, K. (2015). Measuring content diversity in a multi-platform context. *The Political Economy of Communication*, *3*(1), 39–56.

Compaine, B. M., & Smith, E. (2001). Internet radio: A new engine for contact diversity? *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.290293

Correa, T., & Harp, D. (2011). Women matter in newsrooms: How power and critical mass relate to the coverage of the hpv vaccine. *Journalism & Mass*

Communication Quarterly, *88*(2), 301–319. https://doi.org/10.1177/107769901108800205

Culbertson, H. (2007). Agenda diversity indicates newspaper quality. *Newspaper Research Journal*, *28*(1), 40–54. https://doi.org/10.1177/073953290702800104

Dahinden, U. (2002). Biotechnology in Switzerland. *Science Communication*, *24*(2), 184–197. https://doi.org/10.1177/107554702237844

Day, A. G., & Golan, G. (2005). Source and content diversity in op-ed pages: Assessing editorial strategies in thenew york times and the washington post. *Journalism Studies*, *6*(1), 61–71. https://doi.org/10.1080/1461670052000328212

Deutschlandradio. (2019). *Rechtliche Grundlagen - Satzung der Körperschaft des öffentlichen Rechts „Deutschlandradio“*. https://www.deutschlandradio.de/index.media.7ccf879dceda70e621da272e2c230e54.pdf

Deutschlandradio. (2021a). *Deutschlandfunk*. https://www.deutschlandradio.de/deutschlandfunk.2142.de.html

Deutschlandradio. (2021b). *Funkhaus köln*. https://www.deutschlandradio.de/funkhaus-koeln.2089.de.html

Deutschlandradio. (2021c). *Geschichte*. https://www.deutschlandradio.de/geschichte.2091.de.html

Deutschlandradio. (2021d). *Startseite*. https://www.deutschlandradio.de/

Dimmick, J., & Mcdonald, D. G. (2001). Network radio oligopoly, 1926-1956: Rivalrous imitation and program diversity. *Journal of Media Economics*, *14*(4), 197–212. https://doi.org/10.1207/S15327736ME1404_1

Dittmar, J. F. (2011). *Grundlagen der Medienwissenschaft* (2., erw. und korrigierte Aufl.). *Berliner Schriften zur Medienwissenschaft: Vol. 8*. Univ.-Verl. der TU Berlin. http://d-nb.info/1021904937/about/html

Donsbach., W., Wolling., J., Blomberg, C. von, & ilmedia (2015). Repräsentation politischer Positionen im Mediensystem aus der Sicht deutscher und amerikanischer Journalisten. https://www.db-thueringen.de/receive/dbt_mods_00026653

Duckett, J., & Langer, A. I. (2013). Populism versus neoliberalism. *Modern China*, *39*(6), 653–680. https://doi.org/10.1177/0097700413492602

Dvir-Gvirsman, S., Tsfati, Y., & Menchen-Trevino, E. (2016). The extent and nature of ideological selective exposure online: Combining survey responses with actual web log data from the 2013 israeli elections. *New Media & Society*, *18*(5), 857–877. https://doi.org/10.1177/1461444814549041

Eastman, S. T., & Ferguson, D. A. (2013). *Media programming: Strategies and practices* (9. ed.). Wadsworth.

Einstein, M. (2004). The financial interest and syndication rules and changes in program diversity. *Journal of Media Economics*, *17*(1), 1–18. https://doi.org/10.1207/s15327736me1701_1

Elleström, L. (Ed.). (2021). *Beyond Media Borders, Volume 1*. Springer International Publishing. https://doi.org/10.1007/978-3-030-49679-1

Emons, P., Wester, F [Fred], & Scheepers, P [Peer] (2010). "He works outside the home; she drinks coffee and does the dishes" gender roles in fiction programs on dutch television. *Journal of Broadcasting & Electronic Media*, *54*(1), 40–53. https://doi.org/10.1080/08838150903550386

Entman, R. M. (2006). Punctuating the homogeneity of institutionalized news: Abusing prisoners at abu ghraib versus killing civilians at fallujah. *Political Communication*, *23*(2), 215–224. https://doi.org/10.1080/10584600600629844

Entman, R. M., & Wildman, S. S. (1992). Reconciling economic and non-economic perspectives on media policy: Transcending the "marketplace of ideas". *Journal of Communication*, *42*(1), 5–19. https://doi.org/10.1111/j.1460-2466.1992.tb00765.x

Farchy, J. (2008). Promouvoir la diversité culturelle. *Questions De Communication*(13), 171–195. https://doi.org/10.4000/questionsdecommunication.1761

Farchy, J., & Ranaivoson, H. (2011). Do public television channels provide more diversity than private ones? *ENCATC JOURNAL of CULTURAL MANAGEMENT and POLICY*, *1*(1), 50–63.

Federal States of Germany. (2020). *Staatsvertrag zur modernisierung der medienordnung in deutschland*. https://www.rlp.de/fileadmin/rlp-stk/pdf-Dateien/Medienpolitik/ModStV_Text.pdf

Fernández-Quijada, D. (2017). Distinctiveness of public radio in the age of digitization. *Journal of Radio & Audio Media*, *24*(1), 77–89. https://doi.org/10.1080/19376529.2017.1297143

Few, S. (op. 2006). *Information dashboard design: The effective visual communication of data*. O'Reilly Media.

Fletcher, S., & Islam, M. Z. (2018). Comparing sets of patterns with the jaccard index. *Australasian Journal of Information Systems*, *22*. https://doi.org/10.3127/ajis.v22i0.1538

Freedman, E., Fico, F., & Love, B. (2007). Male and female sources in newspaper coverage of male and female candidates in US senate races in 2004. *Journal of Women, Politics & Policy*, *29*(1), 57–76. https://doi.org/10.1300/J501v29n01_04

George, L. M., & Oberholzer-Gee, F. (2011). *Diversity in local television news*. https://docs.fcc.gov/public/attachments/DOC-308602A1.pdf

German Federal States (2020). Medienstaatsvertrag (mstv) vom 14.–28. April 2020.

González-Bailón, S., & Paltoglou, G. (2015). Signals of public opinion in online communication. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 95–107. https://doi.org/10.1177/0002716215569192

Gronemeyer, M. E., & Porath, W. (2014). The ethical demand for editorial diversity in a context of concentrated newspaper ownership in chile. *Palabra Clave - Revista De Comunicación*, *28*(1), 71–101.

Haidar, S., Joly, P., & Chebaro, B. (2005). Style similarity measure for video documents comparison. *Image and Video Retrieval 4th International COnference*, *4*, 307–317.

Haim, M., Graefe, A., & Brosius, H.-B. (2018). Burst of the filter bubble? *Digital Journalism*, *6*(3), 330–343. https://doi.org/10.1080/21670811.2017.1338145

Hanusch, F. (2014). The geography of travel journalism. *International Communication Gazette*, *76*(1), 47–66. https://doi.org/10.1177/1748048513504045

Helberger, N. (2015). Merely facilitating or actively stimulating: Diverse media choices public service media at the crossroad. *International Journal of Communication*, *9*(1), 1324–1340.

Helberger, N., Karppinen, K., & D'Acunto, L. (2018). Exposure diversity as a design principle for recommender systems. *Information, Communication & Society*, *21*(2), 191–207. https://doi.org/10.1080/1369118X.2016.1271900

Hellman, H. (2001). Diversity - an end in itself? *European Journal of Communication*, *16*(2), 181–208. https://doi.org/10.1177/0267323101016002003

Hellman, H., & Vilkko, A. (2019). Is there diversity on popular music radio? Developing a methodology for a quantitative analysis of radio playlists. *Journal of Radio & Audio Media*, 1–19. https://doi.org/10.1080/19376529.2019.1639054

Hendrickx, J., Raats, T., Ranaivoson, H., & Opgenhaffen, M. (2019). 'distinctiveness' and programme diversity in public broadcasting revisited: A seven-country comparison. *International Journal of Media & Cultural Politics*, *15*(3), 283–302. https://doi.org/10.1386/macp_00002_1

Hirschmeier, S., & Beule, V. (2018). Compliance of personalized radio with public-service remits. In T. Mitrovic, J. Zhang, L. Chen, & D. Chin (Eds.), *Adjunct publication of the 26th conference on user modeling, adaptation and personalization* (pp. 15–21). ACM. https://doi.org/10.1145/3213586.3226204

Hirschmeier, S., Beule, V., & Tilly, R. (2020). Translating editorial work into algorithms for personalized radio streams. *Journal of Radio & Audio Media*, 1–22. https://doi.org/10.1080/19376529.2019.1704760

Hirschmeier, S., & Schoder, D. (2020). An approach to explanations for public radio recommendations. In T. Kuflik, I. Torre, R. Burke, & C. Gena (Eds.), *Adjunct publication of the 28th acm conference on user modeling, adaptation and personalization* (pp. 237–240). ACM. https://doi.org/10.1145/3386392.3399306

Hirschmeier, S., Tilly, R., & Beule, V. (2019). Digital transformation of radio broadcasting: An exploratory analysis of challenges and solutions for new digital radio services. In T. Bui (Ed.), *Proceedings of the Annual Hawaii International Conference on System Sciences, Proceedings of the 52nd hawaii international conference on system sciences.* Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2019.602

Hoffman, C., Lutz, C., Meckel, M., & Ranzini, G. (2015). Public service media in the digital age| diversity by choice: Applying a social cognitive perspective to the role of public service media in the digital age. *International Journal of Communication*, *9*(0).

Horwitz, R. B. (2005). On media concentration and the diversity question. *The Information Society*, *21*(3), 181–204. https://doi.org/10.1080/01972240490951908

Huang, H. (2010). Frame-rich, frame-poor: An investigation of the contingent effects of media frame diversity and individual differences on audience frame diversity. *International Journal of Public Opinion Research*, *22*(1), 47–73. https://doi.org/10.1093/ijpor/edp024

Hughes, S., & Prado, P. (2015). Media diversity and social inequality in latin america. In M. Blofield (Ed.), *The great gap* (pp. 109–146). Penn State University Press. https://doi.org/10.5325/j.ctv14gp9cw.8

Humanes, M. L. (2013). *Pluralism and political parallelism in Spanish television news programmes.* https://doi.org/10.4185/RLCS-2013-990en

Humprecht, E., & Esser, F. (2018). Diversity in online news. *Journalism Studies*, *19*(12), 1825–1847. https://doi.org/10.1080/1461670X.2017.1308229

Ibrahimov, O. V., Sethi, I. K., & Dimitrova, N. (2002). Novel similarity-based clustering algorithm for grouping broadcast news. In B. V. Dasarathy (Ed.), *SPIE Proceedings, Data mining and knowledge discovery: Theory, tools, and technology iv* (pp. 294–304). SPIE. https://doi.org/10.1117/12.460239

Iosifidis, P. (2014). Pluralism and public service media. In O. R. Österreichischer Rundfunk (Ed.), *Public value report 2013/14 - texte - it's about the public. It's about the future.* (pp. 51–54). https://zukunft.orf.at/show_content.php?sid=147&pvi_id=1522&pvi_medientyp=t&oti_tag=Texte

Jacobi, C., Kleinen-von Königslöw, K., & Ruigrok, N. (2016). Political news in online and print newspapers. *Digital Journalism*, *4*(6), 723–742. https://doi.org/10.1080/21670811.2015.1087810

Jacobson, S., Myung, E., & Johnson, S. L. (2016). Open media or echo chamber: The use of links in audience discussions on the facebook pages of partisan news organizations. *Information, Communication & Society*, *19*(7), 875–891. https://doi.org/10.1080/1369118X.2015.1064461

Janka, J. (2021, June 30). Interview by F. Terhalle, & S. Hirschmeier.

Jonkman, J. G., Trilling, D., Verhoeven, P., & Vliegenthart, R. (2018). More or less diverse: An assessment of the effect of attention to media salient company types

on media agenda diversity in dutch newspaper coverage between 2007 and 2013. *Journalism*, *19*(11), 1608–1627. https://doi.org/10.1177/1464884916680371

Junge, K. (1994). Diversity of ideas about diversity measurement. *Scandinavian Journal of Psychology*, *35*(1), 16–26.

Juraitè, K. (2014). Recharging public service media discourse: Diversity focus. In O. R. Österreichischer Rundfunk (Ed.), *Public value report 2013/14 - texte - it's about the public. It's about the future.* (pp. 75–78).

Karppinen, K. (2006). Media diversity and the politics of criteria. *Nordicom Review*, *27*(2), 53–68. https://doi.org/10.1515/nor-2017-0230

Karppinen, K. (2007). Making a difference to media pluralism: A critique of the pluralistic consensus in European media policy.

Kiernan, V. (2016). Source diversity among journals cited in science times. *Public Understanding of Science (Bristol, England)*, *25*(2), 198–206. https://doi.org/10.1177/0963662514542908

Kim, D. H., & Kwak, N. (2017). Media diversity policies for the public: Empirical evidence examining exposure diversity and democratic citizenship. *Journal of Broadcasting & Electronic Media*, *61*(4), 682–702. https://doi.org/10.1080/08838151.2017.1375498

Kim, Y., & Jahng, M. R. (2016). Who frames nuclear testing? Understanding frames and news sources in the US and south korean news coverage of north korean nuclear testing. *The Journal of International Communication*, *22*(1), 126–142. https://doi.org/10.1080/13216597.2015.1076732

Kiritoshi, K., & Ma, Q. (2016). Named entity oriented difference analysis of news articles and its application. *IEICE Transactions on Information and Systems*, *E99.D*(4), 906–917. https://doi.org/10.1587/transinf.2015DAP0003

Koeman, J., Peeters, A., & D'Haenens, L. (2007). Diversity monitor 2005. Diversity as a quality aspect of television in the Netherlands. *Communications*, *32*(1). https://doi.org/10.1515/COMMUN.2007.005

Kordus, D. (2014). What's on (digital) tv? Assessing the digital television broadcasting system, its potential and its performance in increasing media content diversity. *Communication Law and Policy*, *19*(1), 55–86. https://doi.org/10.1080/10811680.2014.860830

Kunaver, M., & Požrl, T. (2017). Diversity in recommender systems – a survey. *Knowledge-Based Systems*, *123*, 154–162. https://doi.org/10.1016/j.knosys.2017.02.009

Küng, L., Picard, R. G., & Towse, R. (2009). *The internet and the mass media* (Repr). Sage Publ.

Lacy, S., Fico, F. G., Baldwin, T., Bergan, D., Wildman, S. S., & Zube, P. (2012). Dailies still do 'heavy lifting' in government news, despite cuts. *Newspaper Research Journal*, *33*(2), 23–39. https://doi.org/10.1177/073953291203300203

Lacy, S., Wildman, S. S., Fico, F., Bergan, D., Baldwin, T., & Zube, P. (2013). How radio news uses sources to cover local government news and factors affecting source use. *Journalism & Mass Communication Quarterly*, *90*(3), 457–477. https://doi.org/10.1177/1077699013493790

Lanosga, G., & Martin, J. (2018). Journalists, sources, and policy outcomes: Insights from three-plus decades of investigative reporting contest entries. *Journalism*, *19*(12), 1676–1693. https://doi.org/10.1177/1464884916683555

Lasar, M. (2016). *Radio 2.0: Uploading the first broadcast medium*. Praeger.

Laxton, R. R. (1978). The measure of diversity. *Journal of Theoretical Biology*, *70*(1), 51–67. https://doi.org/10.1016/0022-5193(78)90302-8

Lee, J. K. (2007). The effect of the internet on homogeneity of the media agenda: A test of the fragmentation thesis. *Journalism & Mass Communication Quarterly*, *84*(4), 745–760. https://doi.org/10.1177/107769900708400406

Lee, Y.-C. (2007). Effects of market competition on taiwan newspaper diversity. *Journal of Media Economics*, *20*(2), 139–156. https://doi.org/10.1080/08997760701193761

Levendusky, M. (2013). *How partisan media polarize America*. *Chicago studies in American politics*. The Univ. of Chicago Press.

Li, L., Zheng, L., Yang, F., & Li, T. (2014). Modeling and broadening temporal user interest in personalized news recommendation. *Expert Systems with Applications*, *41*(7), 3168–3177. https://doi.org/10.1016/j.eswa.2013.11.020

Li, Y., & Thorson, E. (2015). Increasing news content and diversity improves revenue. *Newspaper Research Journal*, *36*(4), 382–398. https://doi.org/10.1177/0739532915618408

Lieberson, S. (1969). Measuring population diversity. *American Sociological Review*, *34*(6), 850. https://doi.org/10.2307/2095977

Lister, M. (2009). *New media: A critical introduction* (2. ed.). Routledge.

Loecherbach, F., Moeller, J., Trilling, D., & van Atteveldt, W. (2020). The unified framework of media diversity: A systematic literature review. *Digital Journalism*, *8*(5), 605–642. https://doi.org/10.1080/21670811.2020.1764374

Lõhmus, M., Kõuts, R., Nieminen, H., Kõnno, A., & Aljas, A. (2013). Transformation of newspapers' thematic structure in the 20th century. *Javnost - the Public*, *20*(1), 89–106. https://doi.org/10.1080/13183222.2013.11009110

Lörcher, I., & Taddicken, M. (2017). Discussing climate change online. Topics and perceptions in online climate change communication in different online public arenas. *Journal of Science Communication*, *16*(02), A03. https://doi.org/10.22323/2.16020203

Luhmann, N. (2000). *The reality of the mass media*. *Cultural memory in the present*. Stanford University Press.

Luo, J., Pan, X., & Zhu, X. (2017). Discovery of repost patterns by topic analysis in enterprise social networking. *Aslib Journal of Information Management*, *69*(2), 158–173. https://doi.org/10.1108/AJIM-08-2016-0128

Maekawa, K., Yamazaki, M., Ogiso, T., Maruyama, T., Ogura, H., Kashino, W., Koiso, H., Yamaguchi, M., Tanaka, M., & Den, Y. (2014). Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation*, *48*(2), 345–371. https://doi.org/10.1007/s10579-013-9261-0

Mangani, A., & Tarrini, E. (2018). Social pluralism in public and private television broadcasting. *Javnost - the Public*, *25*(3), 282–297. https://doi.org/10.1080/13183222.2018.1463044

Martin, T. S. (2004). *Broadcast radio programming: A content analysis to measure diversity in radio airplay* [Dissertation]. The University of Alabama, Tuscaloosa, Alabama.

Masini, A., & van Aelst, P. (2017). Actor diversity and viewpoint diversity: Two of a kind? *Communications*, *42*(2). https://doi.org/10.1515/commun-2017-0017

Masini, A., van Aelst, P., Zerback, T., Reinemann, C., Mancini, P., Mazzoni, M., Damiani, M., & Coen, S. (2018). Measuring and explaining the diversity of voices

and viewpoints in the news. *Journalism Studies*, *19*(15), 2324–2343. https://doi.org/10.1080/1461670X.2017.1343650

Matthews, J. (2013). News narratives of terrorism: Assessing source diversity and source use in UK news coverage of alleged islamist plots. *Media, War & Conflict*, *6*(3), 295–310. https://doi.org/10.1177/1750635213505189

May, R. M. (1990). Taxonomy as destiny. *Nature*, *347*(6289), 129–130. https://doi.org/10.1038/347129a0

Mcdonald, D. G., & Dimmick, J. (2003). The conceptualization and measurement of diversity. *Communication Research*, *30*(1), 60–79. https://doi.org/10.1177/0093650202239026

Mcdonald, D. G., & Lin, S.-F. (2004). The effect of new networks on U.S. Television diversity. *Journal of Media Economics*, *17*(2), 105–121. https://doi.org/10.1207/s15327736me1702_3

McQuail, D. (1992). *Media performance: Mass communication in the public interest*. Sage.

Merry, M. (2016). Making friends and enemies on social media: The case of gun policy organizations. *Online Information Review*, *40*(5), 624–642. https://doi.org/10.1108/OIR-10-2015-0333

Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: An empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, *21*(7), 959–977. https://doi.org/10.1080/1369118X.2018.1444076

Moore, M., & Tambini, D. (2018). *Digital Dominance: The Power of Google, Amazon, Facebook, and Apple*. Oxford University Press Incorporated. https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=5359105

Moreau, F., & Peltier, S. (2004). Cultural diversity in the movie industry: A cross-national study. *Journal of Media Economics*, *17*(2), 123–143. https://doi.org/10.1207/s15327736me1702_4

Napoli, P. M. (1999). Deconstructing the diversity principle. *Journal of Communication*, *49*(4), 7–34. https://doi.org/10.1111/j.1460-2466.1999.tb02815.x

Napoli, P. M. (2011). Exposure diversity reconsidered. *Journal of Information Policy*, *1*, 246. https://doi.org/10.5325/jinfopoli.1.2011.0246

Oh, H. J., Hove, T., Paek, H.-J., Lee, B., Lee, H., & Sun Kyu Song (2012). Attention cycles and the h1n1 pandemic: A cross-national study of US and korean newspaper coverage. *Asian Journal of Communication*, *22*(2), 214–232. https://doi.org/10.1080/01292986.2011.642395

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Viking.

Park, D. J., Wang, W., & Pinto, J. (2016). Beyond disaster and risk: Post-fukushima nuclear news in U.S. And german press. *Communication, Culture & Critique*, *9*(3), 417–437. https://doi.org/10.1111/cccr.12119

Park, S. (2005). Competition's effects on programming diversity of different program types. *International Journal on Media Management*, *7*(1-2), 39–54. https://doi.org/10.1080/14241277.2005.9669415

Park, S. (2011). Market change and diversity in the korean movie market. *Asian Journal of Communication*, *21*(6), 544–563. https://doi.org/10.1080/01292986.2011.608224

Park, S.-Y., Yun, G. W., Holody, K., Yoon, K. S., Xie, S., & Lee, S. (2013). Inside the blogosphere: A taxonomy and framing analysis of abortion weblogs. *The Social Science Journal*, *50*(4), 616–624. https://doi.org/10.1016/j.soscij.2013.04.014

Peter, J., & Vreese, C. H. de (2003). Agenda-rich, agenda-poor: A cross-national comparative investigation of nominal and thematic public agenda diversity. *International Journal of Public Opinion Research*, *15*(1), 44–64. https://doi.org/10.1093/ijpor/15.1.44

Pineda, A., & Almiron, N. (2013). Ideology, politics, and opinion journalism: A content analysis of spanish online-only newspapers. *TripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, *11*(2), 558–574. https://doi.org/10.31269/triplec.v11i2.490

Pöchhacker, N., Burkhardt, M., Geipel, A., & Passoth, J.-H. (2017). Interventionen in die produktion algorithmischer öffentlichkeiten: Recommender systeme als herausforderung für öffentlich-rechtliche sendeanstalten. *Kommunikation@gesellschaft*(18).

Powers, M., & Benson, R. (2014). Is the internet homogenizing or diversifying the news? External pluralism in the U.S., danish, and french press. *The International*

Journal of Press/Politics, *19*(2), 246–265. https://doi.org/10.1177/1940161213519680

Raeijmaekers, D., & Maeseele, P. (2015). Media, pluralism and democracy: What's in a name? *Media, Culture & Society*, *37*(7), 1042–1059. https://doi.org/10.1177/0163443715591670

Rodgers, R., Hallock, S., Gennaria, M., & Wei, F. (2004). Two papers in joint operating agreement publish meaningful editorial diversity. *Newspaper Research Journal*, *25*(4), 104–109. https://doi.org/10.1177/073953290402500411

Rodgers, S., Thorson, E., & Antecol, M. (2000). 'Reality' in the st. Louis post-dispatch. *Newspaper Research Journal*, *21*(3), 51–68. https://doi.org/10.1177/073953290002100305

Roggeband, C., & Vliegenthart, R. (2007). Divergent framing: The public debate on migration in the dutch parliament and media, 1995–2004. *West European Politics*, *30*(3), 524–548. https://doi.org/10.1080/01402380701276352

Rohlinger, D., & Proffitt, J. M. (2017). How much does ownership matter? Deliberative discourse in local media coverage of the terri schiavo case. *Journalism*, *18*(10), 1274–1291. https://doi.org/10.1177/1464884916665404

Rohlinger, D. A., Pederson, J., & Valle, G. (2015). Inclusive discourse? Local media coverage of the terri schiavo case. *Sociological Spectrum : The Official Journal of the Mid-South Sociological Association*, *35*(1), 1–25. https://doi.org/10.1080/02732173.2014.978427

Salgado, S., & Nienstedt, H.-W. (2016). Euro crisis and plurality: Does the political orientation of newspapers determine the selection and spin of information? *European Journal of Communication*, *31*(4), 462–478. https://doi.org/10.1177/0267323116659977

Saridou, T., Spyridou, L.-P., & Veglis, A. (2017). Churnalism on the rise? *Digital Journalism*, *5*(8), 1006–1024. https://doi.org/10.1080/21670811.2017.1342209

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2018). What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics.* Advance online publication. https://doi.org/10.1109/TVCG.2018.2864903

Sarrina Li, S.-C., & Chiang, C.-C. (2001). Market competition and programming diversity: A study on the tv market in taiwan. *Journal of Media Economics*, *14*(2), 105–119. https://doi.org/10.1207/S15327736ME1402_04

Schafraad, P., Wester, F [F.], & Scheepers, P [P.] (2013). Media attention to the far-right in three dutch newspapers. *Observatorio*, *2*(7), 1–31.

Schulz, A., Levy, D. A. L., & Nielsen, R. K. (2019). Old, educated, and politically diverse: The audience of public service news. *Reuters Institute Report*.

Scott, D., Chanslor, M., & Dixon, J. (2010). Fair and the pbs newshour: Assessing diversity and elitism in news sourcing. *Communication Quarterly*, *58*(3), 319–340. https://doi.org/10.1080/01463373.2010.503165

Seo, S. (2018). Covering the hermit regime: A comparison of north korea coverage at the associated press and nk news. *Journalism*, *19*(9-10), 1363–1379. https://doi.org/10.1177/1464884918776450

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shumow, M., & Vigon, M. (2016). News diversity and minority audiences. *Journalism Practice*, *10*(1), 52–70. https://doi.org/10.1080/17512786.2015.1014668

Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*(4148), 688. https://doi.org/10.1038/163688a0

Sjøvaag, H. (2014). Homogenisation or differentiation? *Journalism Studies*, *15*(5), 511–521. https://doi.org/10.1080/1461670X.2014.885275

Smyrnaios, N., Marty, E., & Rebillard, F. (2010). Does the long tail apply to online news? A quantitative study of french-speaking news websites. *New Media & Society*, *12*(8), 1244–1261. https://doi.org/10.1177/1461444809360699

Steiner, M., Magin, M., & Stark, B. (2019). Uneasy bedfellows. *Digital Journalism*, *7*(1), 100–123. https://doi.org/10.1080/21670811.2017.1412800

Stirling, A. (1994). Diversity and ignorance in electricity supply investment. *Energy Policy*, *22*(3), 195–216. https://doi.org/10.1016/0301-4215(94)90159-7

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society, Interface*, *4*(15), 707–719. https://doi.org/10.1098/rsif.2007.0213

Stroud, N. J. (2008). Media use and political predispositions: Revisiting the concept of selective exposure. *Political Behavior*, *30*(3), 341–366. https://doi.org/10.1007/s11109-007-9050-9

Sunstein, C. R. (2009). *Republic.Com 2.0* (1. pbk. print). Princeton Univ. Press.

Sweeting, A. (2006). *Too much rock and roll? Station ownership, programming and listenership in the music radio industry*. http://econweb.umd.edu/~sweeting/ASweeting_RadioProgrammingJan06.pdf

Swert, K. de, & Hooghe, M. (2010). When do women gvet a voice? Explaining the presence of female news sources in belgian news broadcasts (2003—5). *European Journal of Communication*, *25*(1), 69–84. https://doi.org/10.1177/0267323109354229

Swert, K. de, & Wouters, R. (2011). The coverage of china in belgian television news: A case study on the impact of foreign correspondents on news content. *Chinese Journal of Communication*, *4*(3), 331–347. https://doi.org/10.1080/17544750.2011.594561

Takens, J., Ruigrok, N., Van Hoof, A. M. J., & Scholten, O. (2010). Old ties from a new(s) perspective: Diversity in the dutch press coverage of the 2006 general election campaign. *Communications*, *35*(4). https://doi.org/10.1515/comm.2010.022

Tan, Y., & Weaver, D. H. (2013). Agenda diversity and agenda setting from 1956 to 2004. *Journalism Studies*, *14*(6), 773–789. https://doi.org/10.1080/1461670X.2012.748516

Thorsen, E., & Jackson, D. (2018). Seven characteristics defining online news formats. *Digital Journalism*, *6*(7), 847–868. https://doi.org/10.1080/21670811.2018.1468722

Toraman, C., & Can, F. (2015). A front-page news-selection algorithm based on topic modelling using raw text. *Journal of Information Science*, *41*(5), 676–685. https://doi.org/10.1177/0165551515589069

Trampuš, M., Fuart, F., Pighin, D., Štajner, T., Berčič, J., Novak, B., Rusu, D., Stopar, L., & Grobelnik, M. (2015). Diversinews: Surfacing diversity in online news. *AI Magazine*, *36*(4), 87–104. https://doi.org/10.1609/aimag.v36i4.2528

Trilling, D., & Schoenbach, K. (2015). Challenging selective exposure. *Digital Journalism*, *3*(2), 140–157. https://doi.org/10.1080/21670811.2014.899749

Turner III, D. W. (2010). Qualitative interview design: A practical guide for novice investigators. *The Qualitative Report*, *15*(3), 754.

Udris, Eisenegger, & Schneider (2016). News coverage about direct-democratic campaigns in a period of structural crisis. *Journal of Information Policy*, *6*, 68. https://doi.org/10.5325/jinfopoli.6.2016.0068

Udris, L., Eisenegger, M., Vogler, D., Schneider, J., & Häuptli, A. (2020). Mapping and explaining media quality: Insights from Switzerland's multilingual media system. *Media and Communication*, *8*(3), 258–269. https://doi.org/10.17645/mac.v8i3.3140

Urban, J., & Schweiger, W. (2014). News quality from the recipients' perspective. *Journalism Studies*, *15*(6), 821–840. https://doi.org/10.1080/1461670X.2013.856670

Vaccari, C., Valeriani, A., Barberá, P., Jost, J. T., Nagler, J., & Tucker, J. A. (2016). Of echo chambers and contrarian clubs: Exposure to political disagreement among german and italian users of twitter. *Social Media + Society*, *2*(3), 205630511666422. https://doi.org/10.1177/2056305116664221

Van der Wurff, R. (2004). Supplying and viewing diversity. *European Journal of Communication*, *19*(2), 215–237. https://doi.org/10.1177/0267323104042911

Van der Wurff, R. (2005). Competition, concentration and diversity in European television markets. *Journal of Cultural Economics*, *29*(4), 249–275. https://doi.org/10.1007/s10824-005-0490-y

Van der Wurff, R., & Van Cuilenburg, J. (2001). Impact of moderate and ruinous competition on diversity: The dutch television market. *Journal of Media Economics*, *14*(4), 213–229. https://doi.org/10.1207/S15327736ME1404_2

Van Hoof, A. M. J., Jacobi, C., Ruigrok, N., & van Atteveldt, W. (2014). Diverse politics, diverse news coverage? A longitudinal study of diversity in dutch

political news during two decades of election campaigns. *European Journal of Communication*, *29*(6), 668–686. https://doi.org/10.1177/0267323114545712

Van Leuven, S., Heinrich, A., & Deprez, A. (2015). Foreign reporting and sourcing practices in the network sphere: A quantitative content analysis of the arab spring in belgian news media. *New Media & Society*, *17*(4), 573–591. https://doi.org/10.1177/1461444813506973

Vergeer, M., Eisinga, R., & Franses, P. H. (2012). Supply and demand effects in television viewing. A time series analysis. *Communications*, *37*(1), 79–98.

Voakes, P. S., Kapfer, J., Kurpius, D., & Chern, D. S.-Y. (1996). Diversity in the news: A conceptual and methodological framework. *Journalism & Mass Communication Quarterly*, *73*(3), 582–593. https://doi.org/10.1177/107769909607300306

Vogel, M., & Arnold, D. P. (2012). *Media of reason: A theory of rationality*. *New directions in critical theory*. Columbia University Press. http://www.jstor.org/stable/10.7312/voge15058 https://doi.org/10.7312/voge15058

Waldhauser, M. (2021, July 1). Interview by F. Terhalle, & S. Hirschmeier.

Webster, J., & Watson, R. (2002). Analyzing the past to prepare for the future: Writing a literature review, *26*.

Welbers, K., van Atteveldt, W., Kleinnijenhuis, J., & Ruigrok, N. (2016). A gatekeeper among gatekeepers. *Journalism Studies*, *19*(3), 315–333. https://doi.org/10.1080/1461670X.2016.1190663

Wirth, T. L. (2001). Nationwide format oligopolies. *Journal of Radio Studies, 8*(2), 249–270. https://doi.org/10.1207/s15506843jrs0802_4

Woods, J. (2007). Democracy and the press: A comparative analysis of pluralism in the international print media. *The Social Science Journal*, *44*(2), 213–230. https://doi.org/10.1016/j.soscij.2007.03.014

Yoon, J. (2013). Who receives media attention in south korea: Analysing internal and external pluralism in the news coverage of policy debates. *Asian Journal of Political Science*, *21*(2), 126–147. https://doi.org/10.1080/02185377.2013.823798

Young, N., & Dugas, E. (2012). Comparing climate change coverage in canadian english and french-language print media: Environmental values, media cultures,

and the narration of global warming. *Canadian Journal of Sociology*, *37*(1), 25–54. https://doi.org/10.29173/cjs9733

Yu, Z., & Zhou, X. (2009). Combining vector space model and category hierarchy model for tv content similarity measure. In *2009 third international conference on multimedia and ubiquitous engineering* (pp. 130–136). IEEE. https://doi.org/10.1109/MUE.2009.33

Zafar, M. B., Bhattacharya, P., Ganguly, N., Gummadi, K. P., & Ghosh, S. (2015). Sampling content from online social networks. *ACM Transactions on the Web*, *9*(3), 1–33. https://doi.org/10.1145/2743023

Zeldes, G. A., Fico, F., & Diddi, A. (2012). Differences in the way broadcast, cable and public tv reporters used women and non-white sources to cover the 2008 presidential race. *Mass Communication and Society*, *15*(6), 831–851. https://doi.org/10.1080/15205436.2011.634084

Zeng, D., Chen, H., Lusch, R., & Li, S.-H. (2010). Social media analytics and intelligence. *IEEE Intelligent Systems*, *25*(6), 13–16. https://doi.org/10.1109/MIS.2010.151

A. Diversity Measurement Overview

Dimension	Measure	Entity	Viewpoint	Topic	Other
Variety	Divide, Count, Report	(Saridou et al., 2017); (Masini & van Aelst, 2017); (Van Leuven et al., 2015); (Bagashka, 2014); (Sjøvaag, 2014); (Young & Dugas, 2012); (Correa & Harp, 2011); (Scott et al., 2010); (Swert & Hooghe, 2010); (Benson, 2009); (Freedman et al., 2007); (Matthews, 2013); (Yoon, 2013); (Armstrong, 2006); (D. J. Park et al., 2016); (Kiernan, 2016); (Koeman et al., 2007);	(D. Rohlinger & Proffitt, 2017); (Masini & van Aelst, 2017) (Merry, 2016); (D. A. Rohlinger et al., 2015); (Bagashka, 2014); (Dvir-Gvirsman et al., 2016); (Baden & Springer, 2014); (Dahinden, 2002); (S.-Y. Park et al., 2013); (Pineda & Almiron, 2013); (Duckett & Langer, 2013); (Schafraad et al., 2013); (R. Rodgers et al., 2004); (Lanosga & Martin, 2018); (D. J. Park et al., 2016); (Lörcher &	(Y. Kim & Jahng, 2016); (Thorsen & Jackson, 2018); (Shumow & Vigon, 2016); (Trilling & Schoenbach, 2015); (Sjøvaag, 2014); (Young & Dugas, 2012); (Swert & Wouters, 2011); (Correa & Harp, 2011); (Smyrnaios et al., 2010); (Pineda & Almiron, 2013); (Yoon, 2013); (Lõhmus et al., 2013); (Udris et al., 2016); (Y. Li & Thorson, 2015); (R. Rodgers et al., 2004); (J. K.	

Dimension	Measure	Entity	Viewpoint	Topic	Other
		(Avraham & First, 2010); (Emons et al., 2010)	Taddicken, 2017); (Roggeband & Vliegenthart, 2007); (Udris et al., 2020)	Lee, 2007); (Lörcher & Taddicken, 2017); (Koeman et al., 2007); (Compaine & Smith, 2001); (Alexander & Cunningham, 2004); (Kordus, 2014); (Udris et al., 2020); (Borum Chattoo et al., 2018)	
	Category Count	(Oh et al., 2012); (Swert & Wouters, 2011); (Lacy et al., 2012); (Lacy et al., 2013); (George & Oberholzer-Gee, 2011)	(Urban & Schweiger, 2014); (Oh et al., 2012)	(Seo, 2018); (Fernández-Quijada, 2017); (Luo et al., 2017); (Toraman & Can, 2015); (Kordus, 2014); (Baum, 2013); (Chambers, 2003); (Berry & Waldfogel, 2001); (George & Oberholzer-Gee, 2011); (Farchy & Ranaivoson, 2011); (Einstein, 2004); (Hendrickx et al., 2019)	

Dimension	Measure	Entity	Viewpoint	Topic	Other
Balance	Simpson Index	(Masini et al., 2018)	(Entman, 2006); (Masini et al., 2018)	(Powers & Benson, 2014); (Vergeer et al., 2012); (Carpenter, 2010); (Tan & Weaver, 2013); (Allen et al., 2017); (Bae, 2000); (Salgado & Nienstedt, 2016); (Y.-C. Lee, 2007); (Mcdonald & Lin, 2004); (Dimmick & Mcdonald, 2001); (Einstein, 2004); (Mcdonald & Dimmick, 2003); (S. Park, 2005); (S. Park, 2011)	
	Shannon Entropy	(Humprecht & Esser, 2018); (Anne Skorkjær Binderkrantz et al., 2017); (Jonkman et al., 2018); (Steiner et al., 2019)	(Humprecht & Esser, 2018); (Jacobi et al., 2016); (Bozdag et al., 2014); (Van Hoof et al., 2014)	(Van Hoof et al., 2014); (Gronemeyer & Porath, 2014); (Takens et al., 2010); (Peter & Vreese, 2003); (Culbertson, 2007); (Möller et al., 2018); (Hellman, 2001); (Sarrina Li &	

Dimension	Measure	Entity	Viewpoint	Topic	Other
				Chiang, 2001); (Steiner et al., 2019); (Aslama et al., 2004); (Farchy & Ranaivoson, 2011); (Mcdonald & Dimmick, 2003); (Hendrickx et al., 2019); (Hellman & Vilkko, 2019)	
	Ordinal Scale Open/Reflec-tive Diversity			(Seo, 2018); (Huang, 2010) (Van der Wurff, 2005); (Van der Wurff & Van Cuilenburg, 2001); (Van der Wurff, 2004)	
	Deviation Index	(Benson et al., 2018); (Hughes & Prado, 2015)		(Benson et al., 2018); (Powers & Benson, 2014)	
	Remaining Measures	(Anne Skorkjaer Binderkrantz, 2012)	(Benson, 2009); (Woods, 2007); (Day & Golan, 2005); (S. Rodgers et al., 2000)	(Hanusch, 2014); (González-Bailón & Paltoglou, 2015); (Humanes, 2013); (Champion, 2015)	

Dimension	Measure	Entity	Viewpoint	Topic	Other
Disparity	Chi-Square Test	(Mangani & Tarrini, 2018)		(Ibrahimov et al., 2002)	
	Cosine Similarity	(Welbers et al., 2016); (Trampuš et al., 2015)		(Trampuš et al., 2015); (L. Li et al., 2014); (Yu & Zhou, 2009)	
	Jaccard Index		(Kiritoshi & Ma, 2016)	(Zafar et al., 2015); (Ibrahimov et al., 2002)	
	Remaining Measures	(Haim et al., 2018); (Sweeting, 2006)	(Jacobson et al., 2016)	(Haim et al., 2018); (An et al., 2014); (Bechmann & Nielbo, 2018); (Möller et al., 2018); (Martin, 2004); (Aslama et al., 2004); (Farchy & Ranaivoson, 2011); (Smyrnaios et al., 2010)	(Haidar et al., 2005)

Table 7: Diversity Measurement Overview

B. Dashboard Mockups

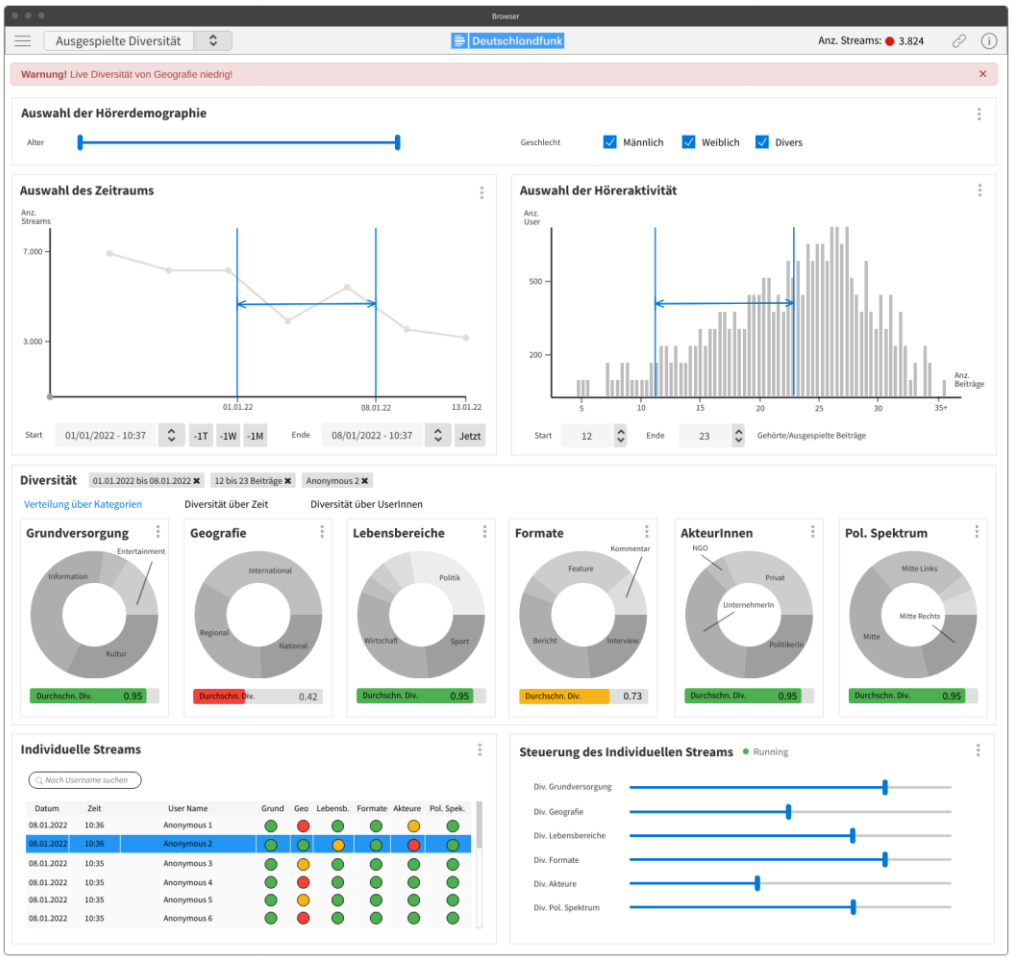

Figure 6: Editors Dashboard - User Filter View

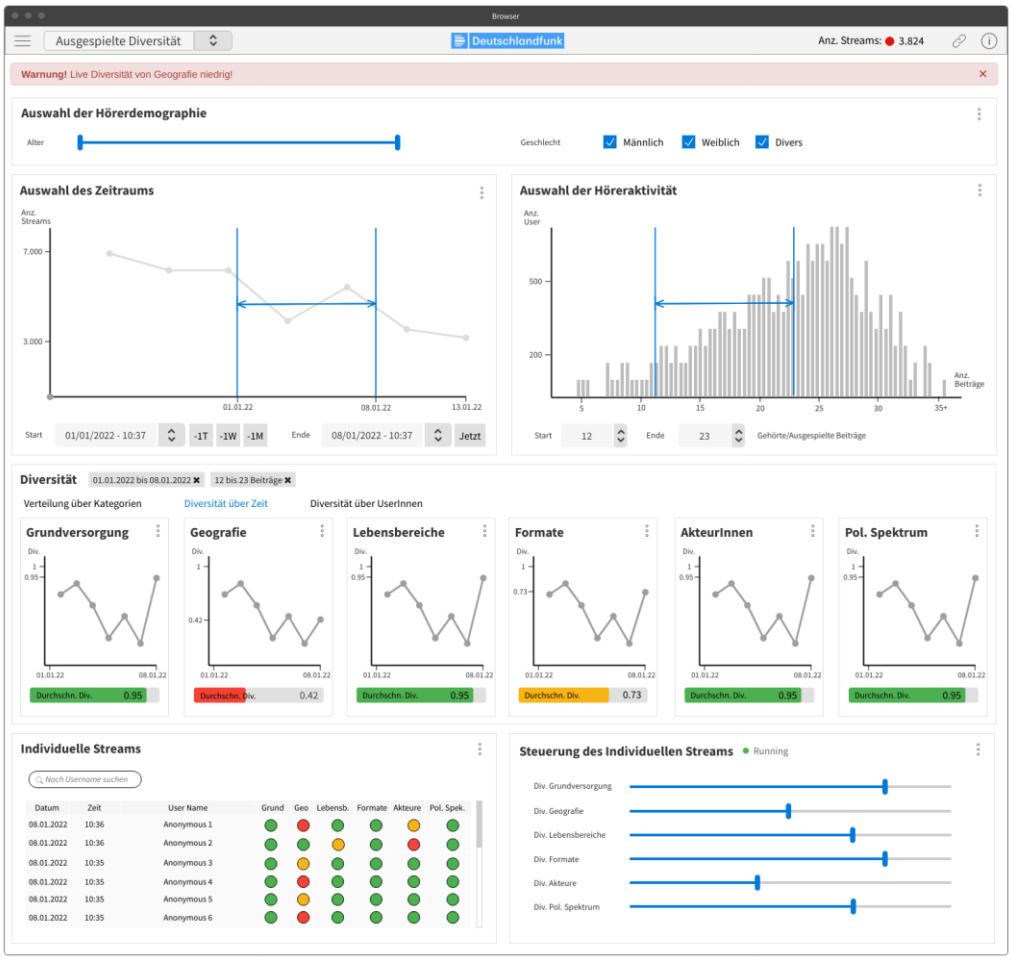

Figure 7: Editors Dashboard - Diversity over Time View

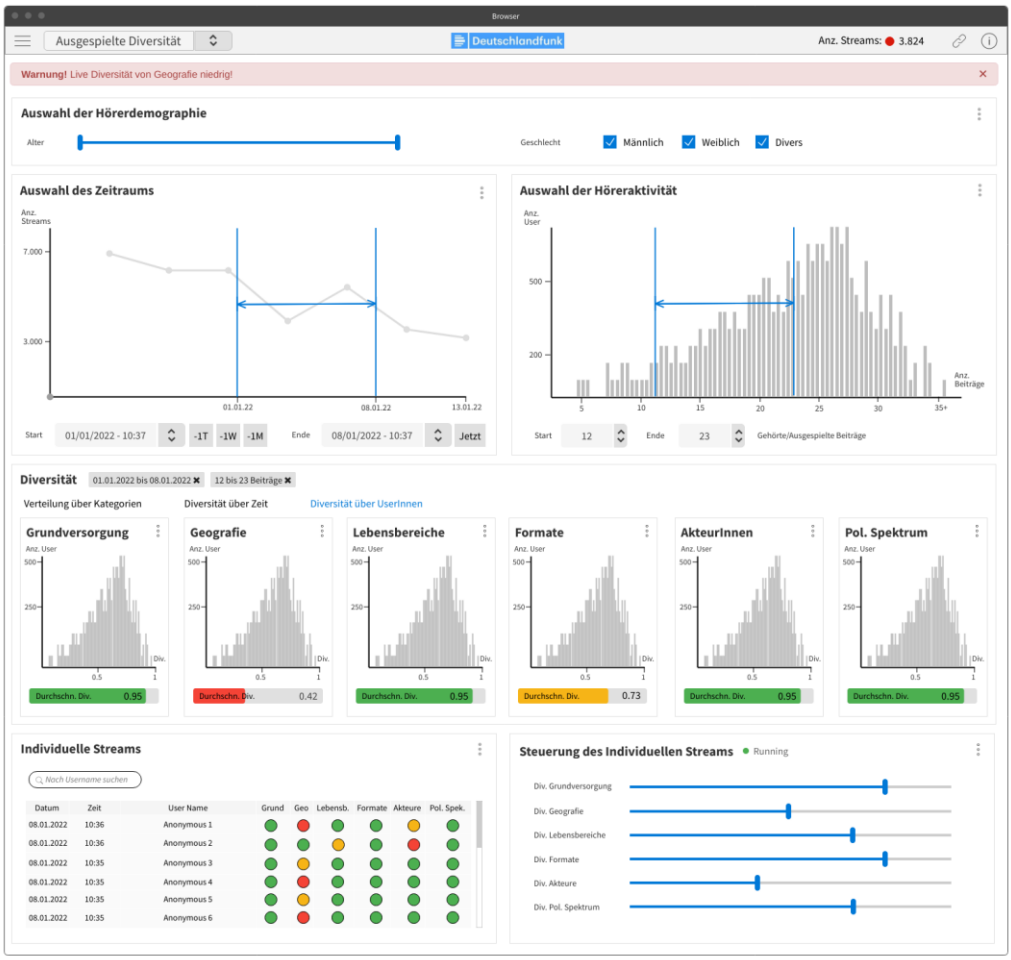

Figure 8: Editors Dashboard - Diversity per User View