HERTIE SCHOOL OF GOVERNANCE

MASTER OF PUBLIC POLICY

CLASS OF 2019

Between Deliberation and Polarization [*]

Analyzing Patterns of Germans' Online Media Diets

Master's Thesis

April 29, 2019

Author:
Moritz Hemmerlein
Hertie School of Governance
mhemmerlein@mpp.hertie-school.org

Supervisor:
Dr. Simon Munzert
Hertie School of Governance
munzert@hertie-school.org

[*]This paper was written as final assessment of the Master of Public Policy program at the Hertie School of Governance. Date of submission was April 29, 2019. The R scripts and some of the raw data are available on Github. For further material and the remainder of the raw data please contact the author.

Acknowledgement

First and foremost, I want to thank those close to me. My mother for her never-ending support and my father for inspiring me to live a life on my own terms. And also my friends, near and far, who encourage me when I lose faith and who have made this whole experience meaningful. I am also grateful to my supervisor Dr. Simon Munzert for his input and advice, as well as the Hertie School of Governance, for providing the space and resources that have enabled me to successfully write this piece of work. And finally, a thank you to all the individuals, past and present, who have laid the foundations for society and academic pursuits to thrive.

Executive Summary

Ongoing disruptions on the market for online media have substantially lowered the costs to consume and create online news content. The Pew Research Center reports that roughly 93% of American adults get at least some news on the internet. This increasing demand was matched by new "born-on-the-web" outlets and the digitalization of "legacy" media that created a high-choice environment enabling people to selectively consume news.

Recent political events have fueled fears that, fostered by demand-driven algorithms on social networks and people's taste for like-minded opinions, the new media reality triggered a sharp increase in polarization. Instead of moderating public discourse and exposing readers to critical and conflicting opinions, people resort to outlets that confirm their beliefs. This, according to sceptics, led to the mergence of echo chambers in which people can avoid cross-cutting content and mainly consume slanted news.

Guess (2018) has challenged this notion and depicts a more differentiated view of news consumption online. Based on browser and survey data of a representative sample of Americans, he shows that across political camps a majority of people consumes mostly news from moderate mainstream sources. Only few people on the extremes of the political spectrum find themselves in filter bubbles and consume ideologically slanted sources.

Hitherto, similar descriptive evidence on the German market of online media remains sparse. I contribute by exploiting recent browser and survey data from 2017 and generate comprehensive insights into patterns of Germans' media diets. I classify domains visited by individuals and extract websites with news-related content. By ranking them on a left-right political alignment scale, I can quantify the slant of people's media diets. Aggregating individual diets across differentials in voting behavior and political leaning, I investigate the common presumption of segregated echo chambers.

The findings are in line with related research and yield compelling evidence that media consumption is largely moderated through mainstream sources and ideologically independent outlets. Only few individuals in the sample who perceive themselves as strictly liberal or conservative exhibit heavily partisan media diets. These findings are robust across multiple methodological specifications and challenge a dominant narrative that online news consumption mostly takes place in isolated filter bubbles.

Contents

List of Figures

List of Tables

List of Abbreviations

2LD	Second-Level-Domain
AfD	Alternative for Germany
BDZV	German Newspaper Publishers Association
CDU	Christian Democratic Union
CSU	Christian Social Union
FDP	Free Democrats
TLD	Top-Level-Domain
PATA	Paying Attention to Attention
OVL	Overlap Coefficient
SLD	Sub-Level-Domain
SPD	Social Democratic Party of Germany
URL	Uniform Resource Locator

1 Introduction

> The newspaper industry (...) serves one of the most vital of all general interests: the dissemination of news from as many different sources, and with as many different facets and colors as is possible.
> U.S. Supreme Court (1944)

The internet age has brought unprecedented dissemination of news from multifaceted sources and substantially lowered the costs to consume them. What has long been seen as vital for democracies, the presence of a media environment that exposes people to a variety of conflicting perspectives, has been turned to a threat in the eyes of many. They envision that, driven by people's preference for opinions that confirm their beliefs and facilitated through demand-driven algorithms, consumers increasingly cocoon themselves in homogeneous echo chambers. Through a series of political events, fueled by misinformation on social media, these fears have materialized and warnings of increased societal polarization became louder.

Despite this widespread perception of the ramifications of internet news, academic research has recently depicted a more differentiated view of the issue. According to this view, the high-choice environment online not only enables homogeneous consumption of news. But through weak ties on social media and the market penetration of mainstream "general-interest intermediaries", exposure to conflicting opinions is far more common than the pessimistic camp suggests. However, the issue is complex and despite potential exposure to cross-cutting content, the understanding for the interdependencies between recent disruptions on how news are consumed, opinion formation and a polarized public remains low.

This research project aims to enrich existing descriptive evidence and quantifies to what extend people consume online news that align with their own political preferences. Recent work by Guess (2018) has approached this question exploiting browser and panel data of a sample of American citizens. In line with prior research, he finds an average overlap of 70% in media diets across political camps. And although he identifies some individuals that consume highly partisan media outlets, the overall results challenge the prevailing perception of widespread selective exposure to ideologically slanted news.

Whilst other studies have come to comparable conclusions, most of recent evidence is on the US market (see Bakshy et al., 2015; Vaccari et al., 2016). The German setting, however, remains hitherto underinvestigated in both, inferential and descriptive terms. This paper sets out to fill these gaps and evaluates most recent data to present comprehensive descriptive insights on the German media landscape. Based on browser and survey data collected from a sample of 1282 German citizens, collected over five month in 2017, I contribute to the controversy on filter bubbles and assess whether Germans' media diets exhibit signs of polarization or are moderate at last.

The data used in this study consists of respondents' browser history coupled with a panel survey on demographics, political preferences and voting behavior. Panelists were incentivized to install passive metering software on their machines that tracked their full browser history. Additionally, respondents were asked repeatedly on their opinions on political events and for their voting decision on September 24, 2017, the day of the last German federal election. Based on this data, I follow the the research design deployed by Guess (2018) and generate insights into the structure of respondents' media diets. I complement the analysis and contrast the overall pattern of media consumption with news that people found on social media and which supposedly were triggered through personalized algorithms and close individual networks.

Based on the browser data observed, I classify and extract domains with genuine news-related content. To account for particular ideological bias of news outlets, I compute an alignment score based on the self-assessed political leaning of their readership. I then turn to the individual level and quantify media diets of individuals according to their overall ideological slant. Aggregating the resulting scores, I finally compare media diets in their extend of overlap across differentials in individuals' voting decisions as well as their political leaning.

In line with Guess (2018), I find remarkable overlap in media diets across political camps. Whilst mainstream media account for the lion's share in media consumption, only few people consume highly partisan news. I accounted for a variety of specifications and preprocessing decisions and found almost exclusively consistent results. Moreover, focusing on news consumption triggered by large social media platforms does not yield significantly different outcomes. Notwithstanding, media diets of heavily partisan individuals exhibit significant structural differences, blurring a fully unambiguous conclusion.

The paper is designed as follows. Chapter two proceeds by discussing related work and embeds the research in a broader conceptualization of the interdependencies between media consumption and polarization. In chapter three, I present the data and discuss the sampling process. Chapter four outlines the preprocessing of the data and the classification process. Chapter five discusses deployed methods and chapter six presents the main results. In chapter seven, I enrich the analysis with further evidence turning from an individual to an outlet perspective. Moreover, I analyze a subset of observations after usage of social media. The last chapter discusses overall results and limitations.

2 Related Literature

The importance of press and mass media for informed and vital democratic citizenry has been widely acknowledged in political science. As the fourth estate, media is supposed to disseminate reliable and independent information as well as expose individuals to alternative, pluralistic opinions that complement narratives of other democratic actors (Dahl, 2008; Nordenstreng, 2000). In his magnum opus, *The Tranformation of The Public Sphere*, philosopher Jürgen Habermas conceptualizes the media as enabler of communication across separated societal groups and facilitator of discourse in the public sphere (Habermas, 1984). As such, it fosters not only public deliberation and informed decision making in a democracy (see Reich, 1988) but serves as a mean to find objective truth. This moderates public debate, prevents polarization and facilitates democratic compromise. Wheather the press can assume this role in the presence of economic and political interests and hold up to the expectations as the watchdog of democracy has been contested (Habermas, 1962; Cage and Goldhammer, 2016). But in any case, if the media affects how people develop opinions and perspectives on the world around them, the extend to which people are selectively exposed to news that propagate only one perspective has a substantial impact on democratic life (Stroud, 2008). Consequently, this conjecture has been thoroughly investigated by political scientists.

Focusing in particular on traditional media such as radio and television, research has shown substantial persuasive and polarizing effects of consuming like-minded news. Dellavigna and Kaplan (2007) present evidence that the accessibility of partisan news substantially increases vote shares for ideologically aligned parties. And recently Martin et al. (2017) fortified this claim by quantifying a feedback mechanism through which people who have preference for slanted news develop polarized opinions which in turn increases their tendency to consume partisan media. Alongside effects on voting behavior, media activity has also been associated with extreme forms of radicalization and violence. As Yanagizawa-Drott (2014) examines, during the Rwandan genocide a total of 10% of the overall violence can be attributed to a popular radio station that encouraged killings against the Tutsi minority. Without being studied scientifically, similar events have already been reported for social media channels, where people were encouraged to violent actions (Steveson, 2018). Anyhow, while traditional, so called "legacy" media channels, have been associated with polarization, evidence on the effects of modern online media environments is ambiguous.

Certainly, digitization driven disruption on the media market continues and the internet increasingly dominates how news are consumed. The Pew Research Center has reported in its annual report on the US news media industry that roughly 93% of American adults get at least some news online. Moreover, even 43% often read their news on mobile devices while the market share of television, radio and print outlets has dramatically declined (Bialik and Matsa, 2017). These developments have increased the accessibility and speed in which information is consumed and induced a tremendous proliferation of new, "born on the web" media outlets (Pew Research Center, 2018).

Induced by these trends, the digital news landscape is regarded as one of "high-choice", where individuals can select among numerous outlets with variations in journalistic quality and differing ideological alignment (Guess, 2018). However, how these conditions materialize in individuals' patterns of media consumption, how they shape individuals' opinions and eventually their actions is complex and subject of academic debate. Stroud (2008) highlights that people can react in various ways, by choosing to avoid political information altogether, seek exclusively congenial information or expose themselves deliberately to conflicting opinions. Regardless of these options of choice, it has become a widespread presumption that, through preferences for confirming viewpoints and people's inertia in muddling through diverse content, the multiplication of news sources online lead to an increase in selective exposure rather than to public moderation (Mutz and Martin, 2001).

The discussion about selective exposure is not new. Almost 60 years ago, Joseph Klapper stated that the tendency of people to select media that are in line with their own beliefs is "widely demonstrated" (Klapper, 1960). Contemporary critics like Cass Sunstein and Eli Pariser have build on his claim and envisioned the internet as a place where filter bubbles prevail and individuals cocoon themselves in echo chambers, reinforcing ideological segregation (Pariser, 2011; Sunstein, 2017). While some research has backed up these fears in more subtle forms (see Prior, 2007), established mainstream newspapers have reported on this notion of the internet and it became a dominating narrative for (actual or perceived) increase in political polarization (Tufekci, 2018; Lobe, 2018; Füchtjohann, 2018). Yet, most recent academic work has depicted a more differentiated reality of media consumption and the political divide of society.

To begin with, research on the ramifications of people's increasing consumption of digital news on political polarization touches upon two distinct questions. First, did the public actually polarize with regard to political ideologies? And second, did people actually resort increasingly to partisan news through proliferation of choice? Even if both is the case, a causal relationship of changing consumption behavior and polarization would still be debatable. However, research on either of the questions has come to ambiguous results.

On the matter of polarization, Lelkes (2016) states that "the debate on mass polarization is itself polarized." In the US context some argued that ideological polarization has dramatically increased among the public as well as political elites (Abramowitz and Saunders, 2005, 2008), while others (Fiorina et al., 2008, 2011; Levendusky and Pope, 2011) attribute those results to errors in statistical measurement and refer themselves to data which interpretation is not unambiguous. Most recent work by Lelkes (2016) aims to reconcile both narratives and concludes, most in line with other related works, that the majority of Americans has not become more polarized. However, partisan individuals with already strong political beliefs have.

Turning to the changes in media consumption, Prior (2013) has acknowledged proliferation on the supply side in terms of an increase in slanted information. His work also states that a less partisan and politically uninterested share of individuals has opted-out from the media market while politically engaged, interested and influential "activists" have immersed themselves into new media sources that match their ideological preferences. From these findings, which are consistent with Lelkes (2016), he concludes that the main threat through increased penetration of the media market by partisan outlets does not lie in polarization of the masses but in a widening gap between partisan activists and the moderate majority of people.

Notwithstanding these analyses of the relationship between digitization of media and polarization, a broad body of work has evolved that focuses in particular on media consumption on social media. In the debate about filter bubbles, platforms like Facebook have been attributed a key role in facilitating the exposure to congenial news through demand-driven algorithms that tend to show people what is in line with their beliefs and preferences. However, a stream of papers since 2015 has built robust insights that imply that exposure to cross-cutting content is much higher than publicly perceived.

To mention a few, Bakshy et al. (2015) analyzed interactions of more than 10 million U.S. Facebook users and find that individuals' exposure to diverse discourse in social media is much higher than it would be "under the digital reality envisioned by some". Vaccari et al. (2016) use a different approach and analyze an online survey mapped to Twitter accounts of respondents. Their findings suggest that homophily, i.e. the existence on homogeneous echo-chambers, is prevalent on social networks. However, they highlight the importance of individual habits and broader patterns of political conversation that determine whether people find themselves in echo-chambers or communities of substantial political disagreement (contrarian clubs). Looking specifically into individual patterns of consumption, Flaxman et al. (2016) uncover that a vast majority of users resort to their favorite, often mainstream, news outlet, which tempers the consequences of the evolvement of a high-choice media environment. However, their overall findings also suggest somewhat more pronounced but modest ideological segregation on social media. All these papers line up with other recent publications that find some selective exposures in consumption of online news, but in absolute magnitude these effects are modest and driven by a small subset of individuals in the population (Barberá, 2015; Barberá et al., 2015; Dubois and Blank, 2018).

Turning to methods, some studies rely on experimental designs, such as Messing and Westwood (2014), who show that the presence of social endorsement decreases the likelihood that individuals share selective partisan content. But the applicability of experiments for large scale analysis of media consumption is limited. Alternatively, many studies resort to classical survey data including scales of self-assessed political alignment to measure polarization over time (e.g Boxell et al., 2017; Lelkes, 2016). However, it has been noted that these scales are often subject to response bias and tend to exaggerate or artificially smooth assessments of political polarization (Hare et al., 2015). In addition, other survey-based measures like self-reported media diets can be biased due to erroneous self-assessment of media consumption and people's tendency to give answers that seem desirable rather than truthful (Guess, 2015).

Cutting-edge studies have thus deployed data that observes individuals in their natural habitat on Facebook or Twitter, where a large share of people's media consumption takes place (Bakshy et al., 2015; Barberá, 2015; Barberá et al., 2015; Eady et al., 2019). These approaches have been promising but are often not representative for the full online population due to demographic factors that correlate with the likelihood to join social media. Moreover, Guess (2018) mentions situational and geographical factors that affect the selectivity of individuals' online networks and hence the exposure to heterogeneous content. He gives the example of an individual who works in finance, who hence has a social network with a taste for a specific type of media content. Focusing on Facebook data would then carry a bias that does not account for news read through other channels and depict an incomplete picture of this person's media diet. Consequently, Guess (2018) analyzes data based on passive metering technology, tracking the full browser history of people, and presents compelling descriptive evidence that media diets of American citizens are widely moderated.

So far, little research has been done that focuses in particular on the German context and descriptive as well as inferential evidence on structural patterns of Germans' media diets is weak. This paper aims to fill some of these gaps and sets out to strengthen the hitherto sparse descriptive evidence on Germans' media consumption patterns across political camps. Based on earlier findings on the extend of selective exposure, I hypothesize that overall media consumption of individuals is largely moderated. However, I expect to find more pronounced polarization on the outmost ends of the political spectrum. Finally, in contrast to the overall consumption of media in people's browsing history, I conduct an analysis based on observations triggered through visits on social media. In this restricted sample, I hypothesize that media diets will increase in polarization but with only small magnitudinal effects.

3 Research Project and Data

This paper draws from prior work and data collected within an overarching research project. The following chapter will briefly outline the project, its methodology and research design and subsequently present the deployed datasets and their collection process in detail.

3.1 Embedding Project: "Paying Attention to Attention"

The embedding project *Paying Attention to Attention: Media Exposure and Opinion Formation in an Age of Information Overload* (PATA) aims to study disruptions on the market for online media and their ramifications on opinion formation in detail (Munzert et al., 2017). The researchers developed a comprehensive research design, capable of answering most pressing questions related to the rise of online media. They explicitly focus on three research questions. First, who is exposed to what kind of information? Second, what are the drivers behind specific patterns of media consumption? And third, how do different characteristics of people play out in the context of online media?

The paper at hand will put a pronounced focus on the first of these questions. I explicitly set out to measure the heterogeneity of media diets across individuals with differing voting behavior as well as ideological political leaning. At this point, it shall be highlighted that I do not claim to isolate causal mechanisms that lead to specific patterns in news consumption. Instead, the main contribution of this paper is the presentation of descriptive insights that point out associations between individuals' characteristics and their tendency to read a particular type of partisan media. Before diving into the applied methods and findings, the following paragraphs outline the data collection process and the structure of the deployed data.

3.2 Sampling Process

All data was gathered via the opinion- and market-research portal *YouGov* from July 1 to December 9, 2017. To ensure representativeness of their panels, *YouGov* follows a sample composition procedure based on weighting and sample-matching (Rivers, 2006). First, a target population is defined, which is the German online population for the data at hand. Based on information from high-quality studies on demographic population marginals (i.e. gender, age and educational attainment), a sampling frame is defined. For the *YouGov* panel, data on marginals from Best for Planning (2017) was used, who conducted 30,000 face-to-face interviews to evaluate the German online population. Finally, a representative stratified sample is drawn from the sampling frame that matches target marginals in the population.

The resulting "target sample" constitutes a representative set of respondents in terms of traditional sampling theory. However, respondents might be hard to contact because they either have never reported their contact details or do not agree to the terms of the survey. Hence, multi-stage matching is applied, combining the representative target sample with *YouGov's* longstanding panel of reliable respondents that contains more than a million members (Munzert et al., 2017). From this panel, a sample of individuals is selected that matches as closely as possible the distribution of the target sample. Through this procedure, *YouGov* guarantees not only a minimum of 1,000 respondents in the survey but also the inclusion of hard-to-reach population subgroups.

3.3 The Pulse Data

At the heart of this analysis lies the Pulse data, a massive dataset collected within the 5-month survey period containing n = 56,102,429 website visits of 1282 panelists. The selected users were incentivized on a opt-in basis to install *YouGov's* proprietary software

Table 1: Panel Participation Over Time

Wave	Participants
1	1516
2	1377
3	1224
4	1215
5	1344

Wakoopa on their devices. The software uses passive metering technology to record people's detailed browser history on a 24/7 basis. Once installed, the full URL path of every website visited, the time and date, and the duration a user stayed on a particular URL was collected. Users could pause tracking for a 15-minutes window and opt-out the panel at any time. Moreover, *Wakoopa* can not distinguish whether domains were actually viewed by users or just open as tabs and the software also registered automatic browser refreshings as additional website visits. *YouGov* encourages their panelists to install the software on all their devices but response rates are lower and tracking technically more limited on mobile devices. Hence, this paper will focus on data tracked on people's desktop machines.[1]

3.4 The Survey Data

The Pulse data was complemented by a survey mapped through an anonymized identifier variable to respondents that installed passive metering on their machines. Over the course of the tracking period, panelists were asked five times to answer questions on their political attitudes, recent political events and their consumption habits of online media. Although the follow-up surveys to the kick-off in July 2017 were pretty short (about 5 minutes), there was some attrition over the first four waves while participation increased again in the fifth wave. The average drop-out rate from wave to wave was -2.6% and there was an average of participation of $N = 1335$ throughout the survey period.

Turning to content, the survey covered a wide range of topics such as people's political preferences, their general attitudes towards politics, their opinion on particular parties and what they think of the election campaign which took place in Germany in 2017. Respondents were also asked for their opinions on specific political issues like Brexit, the refugee crisis or the current economic situation. Moreover, the survey evaluated latent

[1] Not considering news consumption on mobile devices has some shortcomings that will be discussed in section 8.1.

political interest through revealed knowledge of political facts. It also included questions on media usage, social networks (online and offline) and involvement in popular social media services such as Facebook and Twitter.

Central to this analysis is people's self-positioning on a scale of political leaning, ranging from 1 (most liberal) to 11 (most conservative) and their actual vote decision on September 24, 2017, the day of the German federal election. These variables will be used to compute media slant, locate individuals media diets on a left-right scale and examine variations in media diets across individuals with differentials in political leaning and party identification. Details will be provided in section 5. Beforehand, I discuss the deployed classification procedure and preprocessing of the data.

4 Data Preparation

The analysis required some preprocessing and data cleaning. This section discusses the parsing of the raw web domains, the data aggregation procedure and the classification of domains to extract observations with news-related content.

4.1 URL Parsing and Data Cleaning

The raw Pulse data consisted of full URLs that were visited by panelists throughout the tracking period. The data is very sensitive and, although anonymized through numeric de-identification, touches the private sphere of participants. Due to this high sensitivity, I was only granted access to a restricted version of the data, including all observations but with every URL path cut down to the domain level. This comes with some limitations that will be discussed in section 8.1.

Nevertheless, the data contains an exceptional amount of detail including visits of every domain, the date and time it was accessed and the duration an individual remained on that domain. As a first step, I cleaned the data by dropping erroneous observations without an access date or observations that do not comply with the standardized URL scheme and thus does not represent a valid domain. Doing so, I end up with 122,797 unique domains and 55,271,233 visits.

Subsequently, I further cleaned and parsed domains according to their domain host. It has to be highlighted that, although the host of a website is unique, access to the host domain does not necessarily occur through a singular domain but through several sub-level domains (SLD). Taking the example of Facebook, the website is accessible through multiple SLDs, such as *de.facebook.com*, *facebook.de* or *apps.facebook.com*. However, when aggregating the data, for example by summing up the total duration that users consumed a website, domains should be uniquely identifiable with regard to their actual host and content and not be separated on a sub-domain level.

Cutting down all domains to their second-level domain (2LD, host) by removing their top-level domains (TLD) and SLD would render all three domains displayed above to *facebook*. However, not all sub-level domains are supposed to be aggregated on a SLD level. For instance, *mail.google.com* and *news.google.com* are totally different and one does not want to add the duration people stayed on the former to the aggregated duration of the latter.

This issue results in a trade-off in domain parsing. The researcher has to decide either to aggregate data on a SLD level and hence miss aggregating similar websites such as *facebook.de* and *apps.facebook.com*. Or aggregating the data on a higher dimension that confuses websites that have actually different hosts. After detailed examination of the data, I found that this issue is less pronounced for news domains and only introduces a minimum of bias. For this reason, I decided to choose the former approach, cutting domains down to their 2LD before aggregating summary statistics.[2]

4.2 Data Aggregation

After parsing the domains on 2LD level, I end up with 99131 unique domains. I continued by aggregating summary statistics on a domain level, such as average duration a domain was used per day or the total number of visits during the tracking period. The resulting top-domains sorted by average users per day are displayed in figure 1.

[2] I have not conducted an automated, systematic analysis of the implications of different preprocessing steps on the results. This is due to a lack of available computational power to parse and save the Pulse data in various versions and formats. Moreover, restrictions in time and space available to write this paper limit further the level of granularity in which the analysis can be carried out.

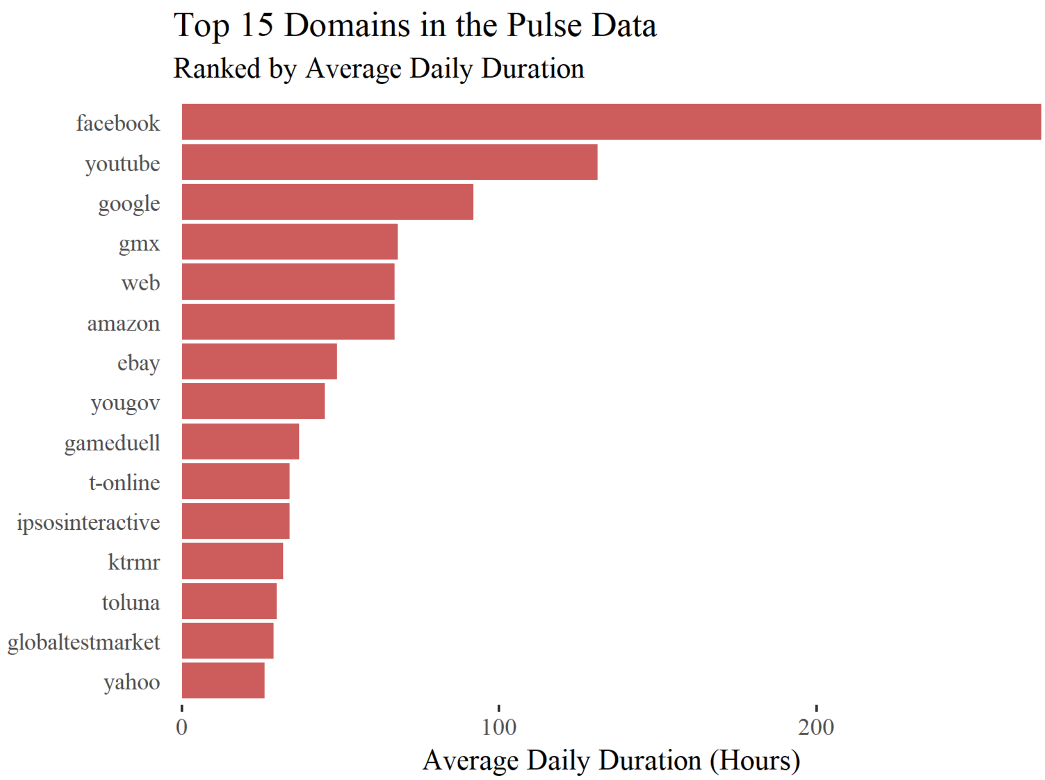


Figure 1: Top Domains

It is notable that Facebook is by far the most used website in terms of duration. This is of particular importance since this analysis does not look into the content people consume directly on social media. When people spend most of their time on this platform, there is a blind spot in this analysis on what people actually consume overall. However, as mentioned in section 2, many studies have put a pronounced focus on Facebook. In contrast, evidence on people's news consumption outside social media remained sparse. It is, thus, a strength of this study to focus on overall news consumption rather than on social media. Nonetheless, section 7.2 will contrast people's overall media consumption and present evidence on a subset of the data that was accessed after visits of social media websites.

A final note on survey websites like *YouGov* and *Ipsos Interactive* among the top domains. It is eminent to the data collection process that many respondents have a tendency to use research platforms either as a hobby or to earn money. However, the representativeness of the sample has been ensured by careful selection of panelists as outlined in section 3.2.

4.3 News Classification

After having aggregated the data on a domain level, I classified domains that are particularly associated with news-related content. Classification was done with a matching process based on predefined dictionaries with labelled domains that were deployed in a lookup algorithm matching labels to domains in the Pulse data.

The algorithm was fed with dictionaries from four different sources. First, I retrieved pre-labelled domain data from www.shallalist.de. The website provides a collection of URLs that are labelled according to predefined categories, such as banking, social media or news. The data is freely available and contains 1.7 millions entries including 40385 news domains from all over the world.

As second source, I deployed data retrieved from www.bdzv.de, the German Newspaper Publishers Association (BDZV). They provide a comprehensive list on the German newspaper landscape, including websites of 698 local outlets as well nationwide-published newspapers. However, this list does not contain online blogs or outlets that are exclusively published online and are not available in print. It does also not include specific magazines and journals that are not registered officially as daily newspapers.

For this reason, I created a third dictionary manually and included outlets that are generally known as discussing political topics such as *Vice*, *Nachdenkseiten*, *The Huffington Post* and *PI-News*. I tried to include all sources that play a relevant role in the German media landscape. However, I cannot rule out that some important outlets are not included.[3]

[3] I tried to pick outlets that cover the whole range of ideological orientation from the far-left to the far-right but some subjective selection-bias cannot be ruled out. The whole manual dictionary is attached in the appendix.

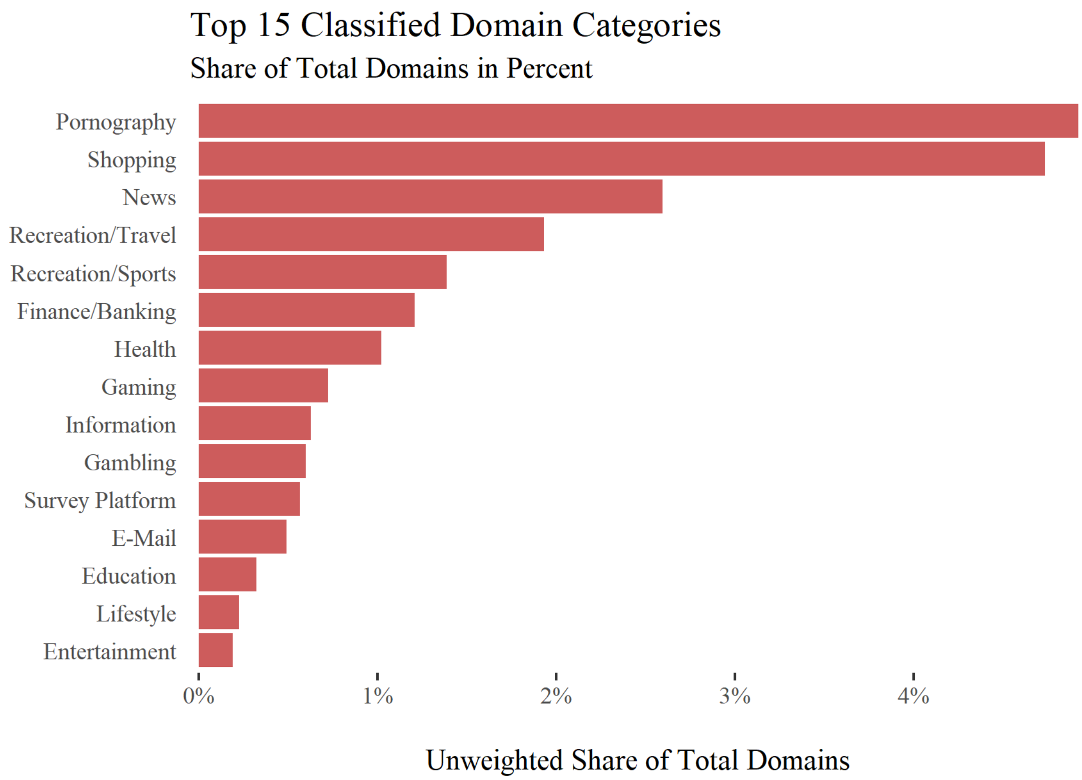


Figure 2: Top Domain Categories

The three lists were eventually given to a lookup algorithm that tested for every domain in the Pulse data whether in was included in one of the dictionaries. If it was, the label attributed to that domain in the dictionary was assigned. With this methodology, I could successfully label 25% of the domains whilst the rest remained unclassified. Moreover, of those domains classified 24.7% were overclassified, i.e they fell into more than one category. If one of these labels was news, I forced the unique domain label to be news and dropped the other labels. Figure 2 displays the 15 top categories that resulted from the classification process.

Of all categories, news domains are the third biggest group with a total of 2572 domains (2.6%). The subsequent section will proceed by describing methods applied to compute ideological slant for all news domains and the construction of individuals' media diets.

5 Media Slant and Media Diets

Once news domains are extracted from the data, I constructed measures for media slant and individuals' media diets. Below, I will outline the used computational metrics as well as discuss subsamples of the data I deployed to construct both scores.

5.1 Estimating Media Slant

Quantifying the political slant of media sources has been a puzzle in political science for a long time. Many studies have assumed that task, but the topic is emotional, methods were often subject to human subjectivity and outcomes contested by many. Moreover, such methods were not only exposed to human bias but also heavily time consuming. This is especially true for traditional techniques that made use of manual counting of partisan phrases, measuring text length of particular parties' news coverage or counting the number of articles addressing specific political issues (Dallmann et al., 2015).

Nonetheless, assessing the neutrality of media sources is an important task, not only for public deliberation and transparency of public discourse. It can also serve as objective benchmark for journalists to assume their obligation to neutral reporting and aid people to choose outlets that hold up to high editorial standards. Recent studies have tried to deploy metrics that are independent from human annotation of textual data. For instance, Groseclose and Milyo (2005) measured media slant by assessing the frequency that media outlets referred to think tanks and compared those with references done by democratic and republican politicians. The more an outlet cited think tanks that were frequently mentioned by liberals the more the outlet was associated with liberal slant and likewise for conservative slant and Republican politicians. Their findings imply a strong liberal bias in the US media landscape and, although not uncontested (see Prior, 2013, p. 33), are regarded as milestone in the quantification of media bias.

Most recently, computational methods have found their way into political science and allow for research designs that deploy extensive large scale text analysis to reliably measure media bias. Gentzkow and Shapiro (2010) quantified slant of 429 local newspapers by selecting phrases from congressional speeches most predictive for party membership. Then, they parsed news articles and counted the frequency that similar phrases were used by an outlet to assess its partisan bias. Calibrating an econometric model of the media market where supply for slanted news is determined by its demand, they find that roughly 20% in variation of media bias can be attributed to the political preferences of its readers. Hence, political leaning of people can serve as a predictor for the slant of outlets they read.

This fact is useful for a methodology applied by Bakshy et al. (2015) who estimate *alignment* scores for shared political news on Facebook. Their method defines a news outlet's alignment score as the average self-assessed political leaning of its readership on a continuous scale. Although these scores are not actual measures of media slant but capture the alignment of outlets to their readers' preferences, they approximate media bias and generate a face-valid alignment of publicly known media sources on a left-right scale (see figure 4).

Additionally, the scores have features that make them valuable for this study. One of these is that the scores provide - in contrast to Dallmann et al. (2015), Groseclose and Milyo (2005) and others - absolute measures for media slant and thus can be computed for individual outlets or even on an URL level. They can also be used to efficiently quantify bias of a large number of outlets, while earlier studies were limited to just a handful of media sources. Moreover, the method yields figures on the same scale on which respondents reported their political leaning. In the case of the German PATA survey this scale ranges from 1 (very liberal) to 11 (very conservative). However, theoretically these scores can be transformed and compared to scores collected on different scales such as the 1-5 scale for the US survey used by Guess (2018).

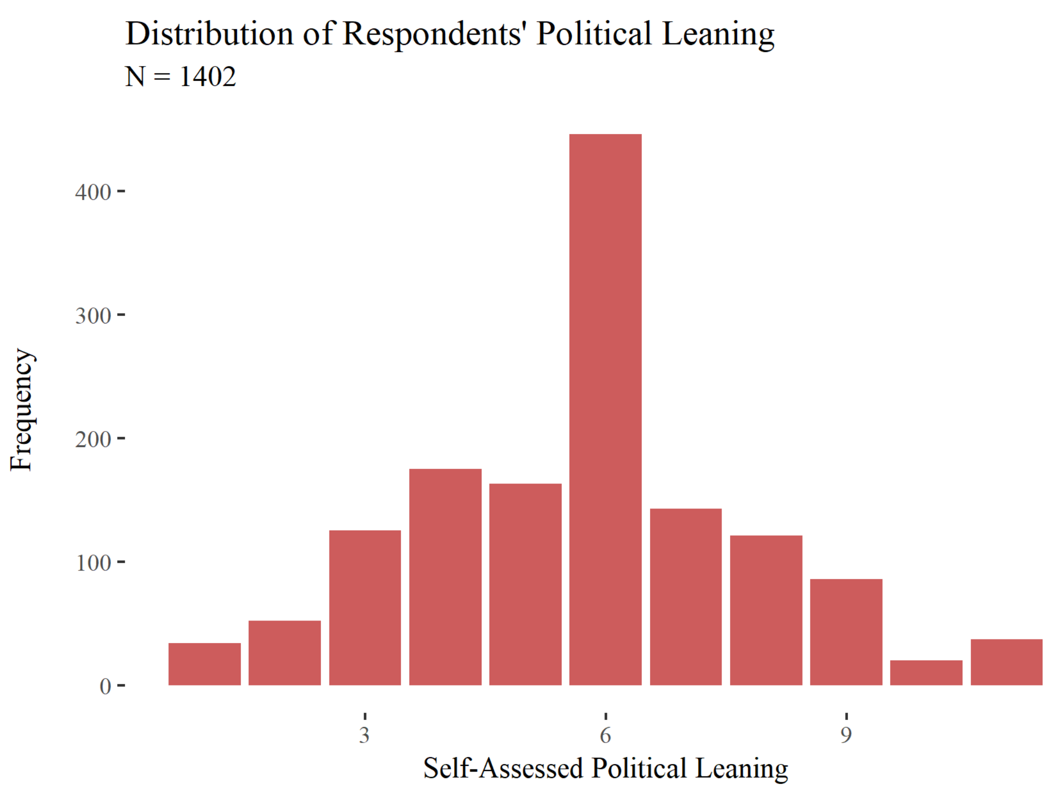


Figure 3: Self-assessed Political Alignment Scores

As one can see from figure 3, the majority of individuals considers themselves as moderate with a political affiliation score of 6. This might be problematic for the analysis since it will potentially align bias of most outlets to the center through the dominance of people with moderate political opinions in the sample. However, although the range of alignment scores does not cover the full range of values on the 1 to 11 scale, there is a substantial spread in outlets' media slant as represented in figure 4. Mathematically, the alignment score is computed with

$$S_i = \frac{1}{n} \sum_{j=1}^{n} o_{ji} \, , \tag{1}$$

where S_i denotes the estimate for an outlet i's slant on a left-right scale. It is calculated on the basis of its readers political opinion o_{ij} with $j \in 1, ..., n$ where every reader gets similar weight. This basic measure does not account for the frequency or duration readers have stayed on an outlet's website. Since some readers visited a particular website just once or rarely and would receive the same weight as readers who consumed a journal frequently, I additionally computed a weighted score that accounts for the total duration a reader has used a domain.[4] Weights are constructed with

$$w_{ij} = \sum_{v=1}^{v} d_{ijv} \, , \tag{2}$$

where d_{ijv} denotes the duration a reader j spent on an outlets website on any particular visit v. The aggregated duration over all visits per reader is then inserted into

$$SW_i = \frac{1}{d} \sum_{j=1}^{n} o_{ji} w_{ij} \, , \tag{3}$$

where SW_i is the duration weighted slant and d is the total duration all readers spent on an particular outlet i's website.

I find, in line with Guess (2018) results on the US market, that the proposed method generates face-valid alignment scores. The unweighted score locates outlets that are generally associated with left ideology like *Der Freitag*, *Nachdenkseiten* and *TAZ* on the left of the scale, while strictly conservative partisan outlets like *Sezession* or *PI-News* obtain very high scores on the scale. However, potentially due to the high number of readers with a moderate 6 on the left-right scale, most of the outlets exhibit very moderate scores.

[4] I also constructed the alignment score weighted by the number of unique page visits but obtained almost similar results.

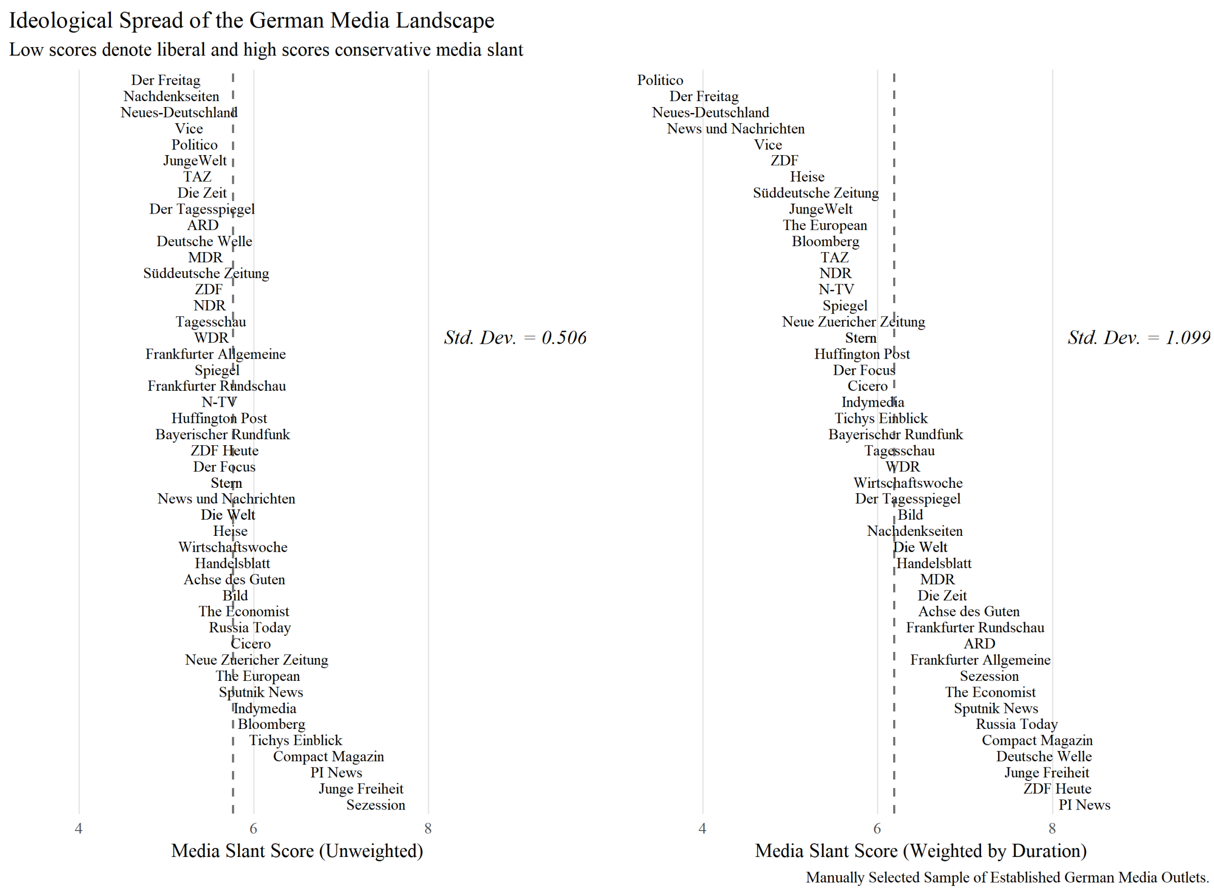


Figure 4: Media Slant

If weighted by duration, the spread of the unweighted scores almost doubles. While most of the partisan outlets are shifted even more towards the extremes, weighting has some unexpected effects. Outlets that are in public discourse associated with rather conservative leaning like *Cicero* shift towards the left, while others like *Deutsche Welle* and *ZDF Heute* move to the far right. The reasons for this are blurry, but most probably related to small sample sizes in outlets' readership in the data that introduce bias in measurement of political slant.[5] Overall, the scores capture alignments of the German media landscape along political camps well. Before presenting the result, it remains to be discussed how from these scores individual media diets are derived.

[5]This issue will be discussed in more detail in section 7.1.

5.2 Estimating Media Diets

Media diets will be evaluated in a two-step process. First, the mean alignment of a person's media diet is evaluated. Second, media diets are aggregated according to individuals' vote decision on the federal election and their political leaning. Both steps will be outlined separately.

5.2.1 Individual Media Diets

On the basis of the slant scores obtained for all news outlets, I turn from the outlet to the individual level and impute the political leaning of an individual's media diet. The term media diet grasps the composition of a person's overall media consumption. It can range from the far left, in case the individual has a propensity to read only news from left aligned media, to the right, if the individual only consumes conservative sources. Both cases would be considered an unbalanced media diet that are unlikely to have much overlap and thus most associated with behavioral patterns predicted in the echo chamber discourse.

An individual's media diet can be expressed as a density function with a location m_i and a spread s_i. Whilst m_i can reach from 1 to 11, like the political affiliation scale, the spread can be zero, if the individual consumes only one outlet. Or, on the other extreme, resemble the spread of the entire media landscape if the individual consumes all media outlets. As with the alignment scores, diets can be constructed as raw measure of overall media consumption or weighted by duration capturing the intensity with which a particular outlet was read.

I will again deploy both, the unweighted method on the basis of the raw alignment scores and the duration weighted diets on the basis of weighted media slant in the analysis. After constructing media diets for individuals, they can be aggregated and compared across different characteristics and studied with regard to the extend of their overlap. The following analysis will focus on the comparison of media diets across people with different party support at the German federal election 2017 and across differentials in political leaning.

5.2.2 The Overlap Coefficient

In order to quantify the overlap in distributions of media diets, I follow Guess (2018) in calculating the Overlap Coefficient (OVL). Initially used by **?** and tested on its statistical robustness by Inman and Bradley (1989), OVL has become a common metric to compare the homogeneity of two populations. It is conceptualized as a "measure of agreement" between two probability distributions and ranges from 0, if both distributions are disjoint and have no point in common, to 1, if both distributions are identical. OVL is defined as

$$OVL = \int_{\infty}^{-\infty} min|f(x), g(x)|\, dX, \tag{4}$$

where $f(x)$ describes probability distribution of the population of interest and g(x) the probability distribution of a benchmark population. Integrating over the joint minima of both functions yields the area shared by both distributions.

The measure has three advantages. First, it is flexible and can be applied to many distributional settings. Second, it provides a commonly understandable metric from 0 to 1 that can be interpreted as fraction or percental overlap. And third, it can be easily transformed in order to adjust to different settings and save computational effort without being transformed in its property to range from 0 to 1 (Inman and Bradley, 1989, p. 3871).

The measure used in Guess (2018) represents an alternative way of computing OVL and has recently been used in related publications aiming to quantify public polarization (see Levendusky and Pope, 2011; Lelkes, 2016). It is based on the dissimilarity index capturing the probability mass under two distributions that is not shared with the other. Since the dissimilarity index also ranges from 0 to 1, but with inverse implication, it can be subtracted from 1 to arrive at OVL. The measure is calculated as

$$OVL = 1 - \frac{1}{2} \int_{\infty}^{-\infty} |f(x) - g(x)|\, dX \tag{5}$$

where $f(x)$ and $g(x)$ again denote the probability distributions of two populations while the interpretation of OVL does not change. This formulation is useful, since in the analysis at hand the exact densities of media diets are unknown. In such setting, Schmid and Schmidt (2006) show that OVL can be approximated based on non-parametric kernel density estimates for $f(x)$ and $g(x)$. I have deployed this method and computed kernel densities for individuals' media diets aggregated by voting decision at the federal election. I end up with aggregated distributions for supporters of a particular party.

In addition, although voting and political leaning are correlated, I used the same method to compare media diets across groups with different political affiliation. I labelled people with values below 4 on the left-right scale with "liberal", people with a score above 8 as "conservative" and the remaining individuals as "moderate". Then I calculated the kernel density distribution of media diets for all three groups respectively.

Once I arrived at reliable kernel density estimates for all subgroups, I compared every group's aggregated media diet with the population average by computing OVL scores. Moreover, I compared all pairs of subgroups with each other, to not only get estimates for groups' media diets in contrast to the overall population but to compare groups directly. Results are reported in section 6.[6]

What is more, OVL also comes with some downsides that shall be discussed. One issue is that OVL does not make a statement about the location of the common probability mass shared by two groups (see Gastwirth, 1975). This would be in particular of interest for the research at hand, since additional to comparing media diets of two subgroups with regard to their overlap, it would be interesting to see their common location on the political alignment scale. Moreover, as outlined in Clemons and Bradley (2000), when similarities of two distributions increase, the bias of OVL as an estimator of overlap increases. This bias is minimal, but should be kept in mind since obtained kernel density estimates for media diets are highly similar across groups and thus their OVL might be subject to some bias.

A last issue of the outlined methodology is the reciprocity in construction of media alignment scores and individuals' media diets. While media slant is computed on the basis of readers' political leaning, users' media diets are constructed as average alignment scores of outlets they read. This induces a feedback-loop that shifts the ratings of an individual's media diet towards his political leaning. However, this mechanism would be expected to drive media diets of people with different political leaning further apart, an expectation that is not consistent with the findings as they will be presented in the next chapter.

5.3 Sample Selection

One further methodological issue is concerned with sample selection. The fine-grained structure of the data gives a lot of leeway to the researcher in preprocessing decisions. Some of these have already been outlined in previous sections, like parsing of domains, aggregation and weighting. However, after classifying and filtering domains associated with news from the raw data, there remains a large list of websites that are not equally important for tracking individuals' consumption of partisan content. Among sources that

[6] I will not report confidence intervals for OVLs as proposed by Inman and Bradley (1989) nor test the significance of differences in overlap coefficients across subgroups. However, this is an interesting topic for further research that aims to dig deeper into the structural differences between individuals with regard to their media diets.

mainly contain political content, such as *cicero.de* or *sueddeutsche.de*, there are others which cover either a specific field like *computerbild.de* or a large spectrum of information such as web portals like *gmx.de* without a particular focus on politics or news.

Related studies such as Bakshy et al. (2015) as well as Guess (2018) have the advantage to work on an URL level. For this reason, they can introduce a distinction between "soft" and "hard" news. They use machine learning classifiers to label content of particular URLs with regards to their content as either genuinely presenting political hard news or a wider range of news related content not specifically political as soft news. Bakshy et al. (2015) report substantial bimodality in alignment scores that is commonly associated with polarized distributions for hard news. Soft news, on the contrary are distributed in a unimodal shape.

Since this study works with data on the domain level, I could not apply machine learning to train a hard news classifier to distinguish observations based on the their content. Hence, I took an alternative approach by running the analysis on increasingly restrictive subsets of the data, incrementally dropping domains that are not associated specifically with political news content. This way, I aim to simulate an analysis on hard news content, since dropping apolitical domains, like sports news, would increase the weight of political content in the data the more restrictive my samples get.

As a first step, computed slant scores and media diets on the basis of the full list of outlets. From outlined theory and previous findings one would expect that in particular partisan outlets with a focus on politics have strong associated alignment scores. The full sample that includes relatively less of such partisan sources is thus expected to show only modest patterns of partisan segregation due to inclusion of apolitical and non-partisan content moderating overall media diets. Thus it will serve as a benchmark for more restrictive subsamples that, if the theory holds, are expected to show more diverging patterns in aggregated media diets compared to the full sample. In total I created two restrictive samples as follows.

First, I extracted a subset of the full sample of major German news outlets as issued by the list of BDZV newspapers which was introduced in section 4.3. This way, I got rid of large web portals, magazines, sports websites or entertainment platforms that provide none or little content related to politics and restricts my sample to newspapers that all report at least to some extend on political topics. As table 2 shows, the sample of major German outlets captures 8.02% of total visits and 12.70% of total duration spend on news domains and consists of 231 outlets.

Lastly, I manually selected the biggest and most relevant outlets in public discourse that all have a primary focus on political news (all displayed in figure 4. Selection was done subjectively, but I aimed to include both, mainstream outlets with huge share of overall traffic as well as smaller independent outlets or with strong partisan image. I balanced outlets with leftist and conservative leaning and restricted the selection to outlets with at least 10 average views or 5 minutes of average consumption per day. The goal was to get a balanced sample of news outlets that focus on politics and account for a large share of overall traffic. In sum, the restricted sample of only 44 websites account for 19.24% of total duration spend on news websites an 12.5% of total visits. Taking into account that the benchmark sample contains traffic for sports and some entertaining websites as well, the restricted sample reasonably grasps a substantial amount of politically focused media consumption by individuals in the data.

Table 2: Percentage of News Traffic by Sample

Sample	n	Visits (in 1000)	Duration (in hours)	% Total Visits	% Total Duration
All Outlets	2574	5447.96	50518.86	100%	100%
Restricted Outlets	271	775.25	10728.94	14.23%	21.23%
Selected Outlets	46	678.09	9753.73	12.44%	19.30%

Through the stage-wise subsetting process I hope to get a complete panorama of evidence that is supported by the data. While the full sample is most exhaustive in describing patterns of overall news consumption without a focus particularly on politics, websites with foremost political content gain incrementally in weight by applying the restricted subsetting process. I hypothesize that due to increased weight of traffic on political websites, weight of partisan traffic increases as well. For this reason, I expect the distribution of media diets to become more bimodal with an increase in sample-restrictiveness. I will now proceed to evaluate media diets across all metric specifications and subsamples (see table 2) as they were presented in this chapter.

6 Evaluating Media Diets

I will now turn to the presentation of the main results. All findings will be presented for all six previously discussed specifications, i.e. for weighted and unweighted alignment scores along three different subsamples of the data. All six variants of the analysis will be presented graphically and in terms of numeric figures. The last part of this chapter evaluates specific determinants of people's media diets using ordinary least squares regression.

6.1 Distribution of Media Diets

The constructed densities in media diets can be evaluated graphically, which I will do first, and quantified in their extend of overlap using the presented OVL scores.

6.1.1 Graphical Representation

Figure 5 displays the distributions of the weighted and unweighted media scores for all three subsamples of the news data. At a first glance, these distributions show a remarkably homogeneous picture of people's media consumption across all specifications and subsamples. This will remain the comprehensive impression throughout all analyses conducted. This finding is in line with most of the recent literature on the homogeneity of online media consumption and in particular consistent with the results in Guess (2018). However, some peculiarities in the observations require detailed interpretation and allow for a differentiated conclusion.

Comparing the overall differences in spread between the weighted and unweighted scores shows that across all samples the spread for the unweighted scores is much lower, a pattern I already pointed out in section 5. Since the unweighted scores count visits of all individuals with equal weight, no matter how long they stayed on a website, these scores yield insights into which websites are known to whom. If some groups show substantial underrepresentation in some range of the alignment score that means that many group members have almost never accessed websites that are aligned accordingly on that location of the scale.

This is in particular interesting for *AfD* and *Linke* voters, who show a large spread on the unweighted score compared to other parties (see figure 5 a). While *AfD* voters are highly underrepresented left of the population median (indicated by the white line), *Linke* voters rarely exhibit media diets on the right of the 50%-quantile. This pattern vanishes for the duration weighted score b, however, it indicates that the spread of news websites that are at least visited once among these subgroups is larger than for others. This might indicate a bigger motivation to look beyond the mainstream for other news sources.

As mentioned, the pattern is not as pronounced for the weighted scores. This is a hint that, although many other sources are tried out, the majority of voters across all parties

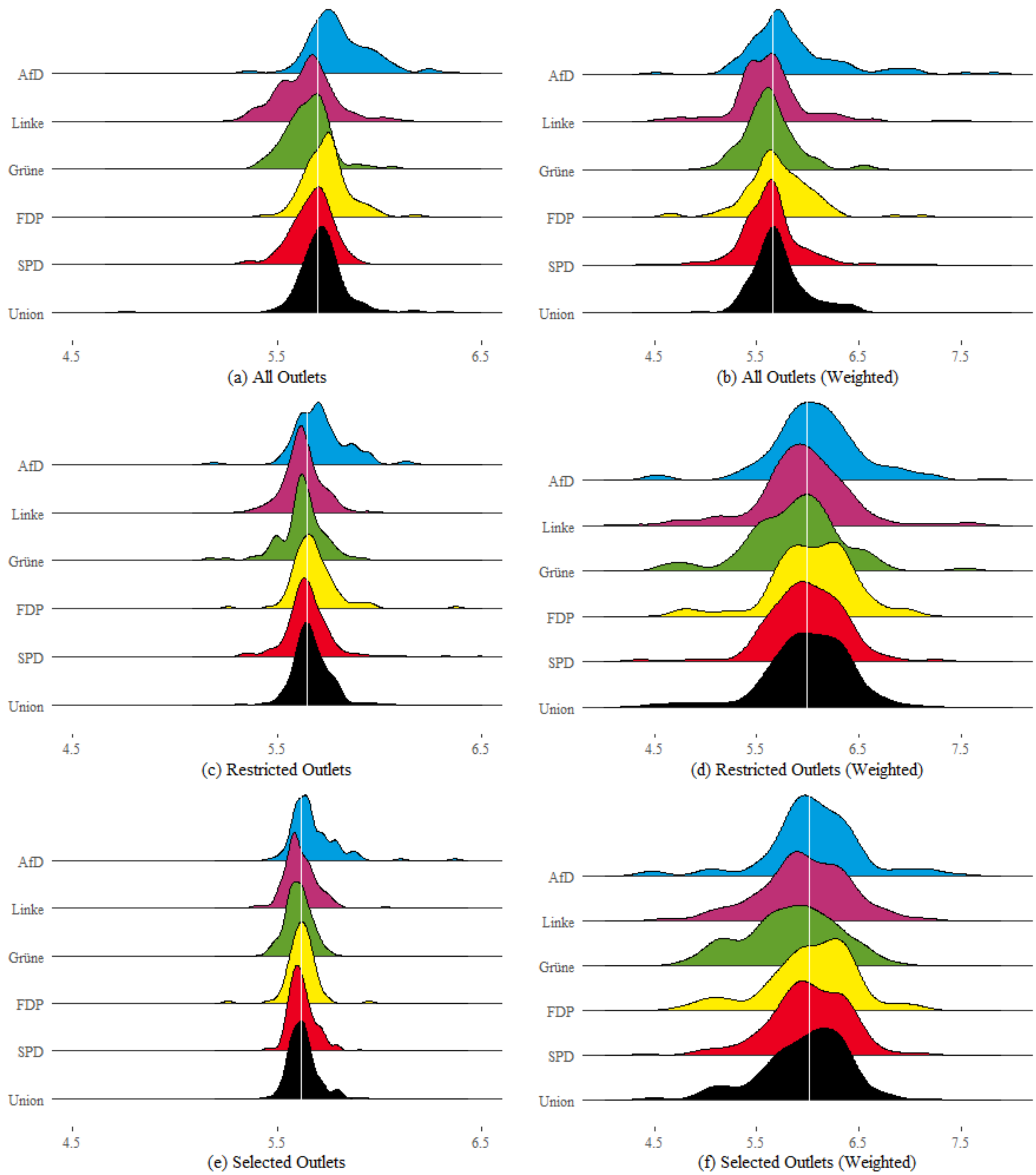


Figure 5: Aggregated Media Diets by Party Support

consumes most of the time media outlets in the center of the alignment scale. But yet, even in the duration weighted scores, there is a long but modest tail to left or right respectively for *Linke* and *AfD* voters that shows that even measured in terms of duration there is a subgroup of people who consumes sources that are slanted towards the extremes of the political spectrum. Nonetheless, this group is almost negligible compared to the majority of people in the data.

Instead of grouping by vote decision, I grouped individuals on the far left and right of the self-assessed alignment scale to be "liberal", if they reported a score lower than 4, or "conservative", if they reported a score higher than 8. Taking this ternary classification as grouping measure shows that the outmost ends of the media landscape are almost exclusively consumed by individuals who perceive themselves as very liberal or very conservative.

Again, the sample including all outlets shows distributions that become increasingly disjoint for liberals and conservatives, while the density of moderates almost mimics a normal distribution (see figure 6 a and c). Weighting the full sample by duration shows that, again, the patterns get more modest. However, there is still substantial degree in left-aligned media diets for liberals as well a pronounced "bump" to the right for conservatives.

This pattern remains in the restricted sample of major German news outlets but vanishes for the selected sample (see e). Here, the distributions for unweighted scores are almost identical across vote decision as well as across political leaning. This is not surprising because this sample contains mostly well known media outlets in Germany and it is very likely that respondents visited at least once during the survey. Nonetheless, there is a higher relative share of *AfD* voters on the right which indicates that some sources are only visited by them. In contrast, such patterns do not exist for other parties in the unweighted graph. In the weighted manual selection I find disjoint media diets for a small group of moderates and liberals compared with conservatives (bumps on the right and left of figure 6 f). But as pointed out earlier, this finding is marginalized by the overwhelming majority of people that consumes media diets centered around the population median.

6.1.2 Numeric Analysis

Enriching the graphs with some objective quantification, table 3 shows the mean in media diets for all parties across all data subsets. As previously suggested by the graphs, the means of all subgroups are exclusively centered around the population means (last

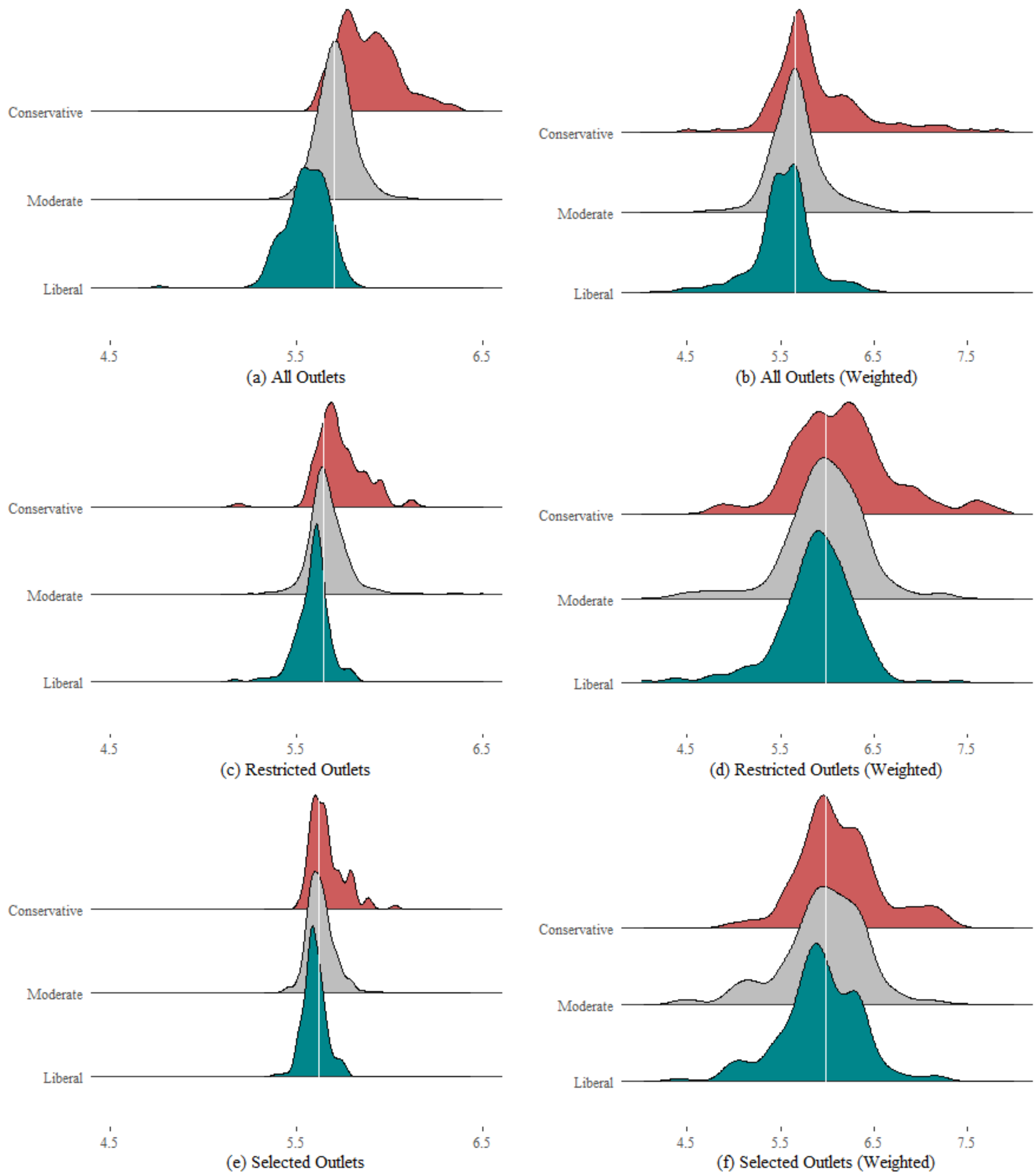


Figure 6: Aggregated Media Diets by Political Leaning

column). Technically, none of the groups aggregated by party support exhibits an average further away from the pooled mean than one standard deviation. This holds for all subsets of the news domains and the weighted and unweighted scores respectively.

Table 3: Mean and SD in Media Diets

	Union	**SPD**	**Grüne**	**Linke**	**FDP**	**AfD**	**Other**	**Pooled**
All Outlets	5.72 (0.13)	5.68 (0.12)	5.65 (0.11)	5.64 (0.14)	5.74 (0.11)	5.8 (0.15)	5.7 (0.11)	5.7 (0.14)
All Outlets (Weighted)	5.72 (0.26)	5.61 (0.34)	5.63 (0.26)	5.61 (0.41)	5.71 (0.35)	5.87 (0.55)	5.6 (0.29)	5.68 (0.38)
Restricted Outlets	5.67 (0.09)	5.65 (0.13)	5.62 (0.12)	5.63 (0.12)	5.69 (0.18)	5.71 (0.13)	5.67 (0.08)	5.66 (0.13)
Restricted Outlets (Weighted)	6.02 (0.45)	5.96 (0.56)	5.88 (0.47)	5.93 (0.52)	6.07 (0.46)	6.12 (0.6)	5.86 (0.54)	5.99 (0.52)
Selected Outlets	5.62 (0.07)	5.62 (0.06)	5.6 (0.06)	5.61 (0.08)	5.62 (0.07)	5.67 (0.11)	5.65 (0.08)	5.63 (0.08)
Selected Outlets (Weighted)	5.96 (0.43)	6.01 (0.39)	5.83 (0.45)	5.97 (0.44)	6.02 (0.45)	6.06 (0.48)	5.91 (0.39)	5.98 (0.44)

The table displays the mean and standard deviation (in brackets) in media diets broken down by party support. The last column displays the population mean and standard deviations. Statistics are displayed for all subsets of news based on which diets were constructed.

The comparison of means, however, does not make claims about the overall extend in overlap. However, the OVL scores also do not allow for a different conclusion. Table 4 displays the overlap in media diets for all parties across all variants of analysis. Most remarkably is the low overlap score of 70% for *AfD* voters in the full sample which remains at "only" 81% when weighted by duration. This suggests that 19% of people who vote for the *AfD* exhibit media diets more biased towards the political right than the aggregated population.[7] Nevertheless, this finding is not very robust to different subsets of news. Considering the selected sample, OVL increases up to 90% across all examined groups of party identification.

I also compared groups directly in their extend of media diet overlap. In this configuration, I only included the unweighted score for all outlets and displayed the weighted scores for the other news subsets exclusively. Figure 7 shows that the overlap between supporters of the *AfD* and *Grüne* is below 75% for four of the five specifications. Only the subsample of major outlets shows an OVL of 83%, which still implies a fraction of 17% of people that have structurally diverging media diets.

[7] The score itself does not make a claim of the direction of the deviation between two distributions. However, in combination with the graphical impression, this conclusion is reasonable.

Table 4: Overlap Scores

	Union	SPD	FDP	Grüne	Linke	AfD
All Outlets	0.877	0.906	0.835	0.810	0.794	0.701
All Outlets (Weighted)	0.892	0.917	0.853	0.890	0.881	0.810
Restricted Outlets	0.908	0.941	0.893	0.872	0.839	0.780
Restricted Outlets (Weighted)	0.938	0.947	0.887	0.863	0.931	0.897
Selected Outlets	0.946	0.919	0.883	0.890	0.891	0.819
Selected Outlets (Weighted)	0.944	0.955	0.910	0.827	0.936	0.899

The table displays the the overlap of party level kernel density estimates in media diets with the population. It is interpreted as fraction of the population media diet and ranges from 0 (totally distinct) to 1 (identical). The overall mean is 0.88.

Finally, I did a similar analysis for the ternary indicator variable of respondents political leaning. As figure 8 shows, the overlap for all outlets between liberals and conservatives is very low with only 22%. With moderates both partisan camps do not share more than 55% in distribution of media diets for the sample including all outlets. These figures, however, also turn out to be not very robust and increasingly converge towards an overlap between 86% and 92% across political camps.

6.2 Determinants of Media Diets

So far, media diets have only been investigated on an aggregate level across party identification and political leaning. To broaden the overall picture on what determines individual media diets, I follow Guess (2018) and regress the weighted and unweighted media diets on demographics, party affiliation or political interest. I will use a standard OLS specification as follows.

$$S_i = \alpha + \beta_j v_{ij} + \gamma X_i + \epsilon_i \,, \tag{6}$$

The formula represents a regression model with an individual's media slant score S_i as independent variable. v_{ij} depicts a binary variable whether and individual i voted for party j, where $j_i \in J = \{Union,\ SPD,\ FDP,\ Gruene,\ Linke,\ AfD\}$ and $\sum_j^J v_{ij} = 1,\ \forall_i$. *Union* voters represent the benchmark category. Moreover, I included control variables into the model denoted by the control matrix X_i. X includes information on gender, age, educational attainment, household income and political interest. The variables are coded as follows, whereas the first category always represents the reference category for interpretation.

Gender is coded as *Gender = {Male, Female}*. Age is coded as *Age = {< 30, 30-39, 40-49,*

Figure 7: Overlap In Media Diets By Party

Figure 8: Overlap In Media Diets By Political Leaning

50-59, >60}. Educational attainment is coded as *EA = {No Higher Education, Apprenticeship, College}*. Interest in politics is coded as *Political Interest = {No Interest, Slight Interest, Moderate Interest, Interest, High Interest}*. All results are displayed in table 5.[8]

The findings generally confirm the picture depicted so far. However, some additional interesting insights can be noted. First of all, being a voter of the *AfD*, *Grüne*, *SPD* or *Linke* is indeed associated with diverging patterns in media diets compared to *Union* voters. Considering column 1, the specification with unweighted media diets on the basis of all outlets as independent variable, shows that the coefficients for these three parties are significant on the highest of three conventional levels. For the *FDP*, the effect remains insignificant which mirrors a larger ideological alignment of *FDP* and *Union* voters. In terms of practical significance, however, all effects are only modest with about a half standard deviation shift of media diets (Std. Dev. = 0.136).

The effect for *AfD* voters is positive and denotes a conservative slant that gets more pronounced in absolute terms in the weighted specification. But relatively, the effect remains low with a shift of half a standard deviation compared to *Union* voters. *Grüne*, *SPD* and *Linke* voters show effects in the opposite direction but the effects are even more modest. Overall, none of the coefficients is very robust whilst all loose in significance for the regression models applied to the restricted subsets of news.

[8] I also estimated the model including the ternary categorical variable of people's political leaning instead of party identification. The results are attached in the appendix.

Table 5: Determinant Analysis Across Specifications

	All Outlets Raw	All Outlets Weighted	Restricted Outlets Raw	Restricted Outlets Weighted	Selected Outlets Raw	Selected Outlets Weighted
	(1)	(2)	(3)	(4)	(5)	(6)
Age: 30-39	0.03* (0.01)	0.03 (0.04)	0.03** (0.02)	−0.09 (0.06)	0.04*** (0.01)	−0.004 (0.06)
Age: 40-49	0.02 (0.01)	0.03 (0.04)	0.01 (0.02)	−0.10 (0.06)	0.02** (0.01)	0.01 (0.05)
Age: 50-59	0.02* (0.01)	0.04 (0.04)	0.01 (0.01)	−0.04 (0.06)	0.03*** (0.01)	0.02 (0.05)
Age: >60	0.02 (0.01)	0.04 (0.04)	0.02 (0.02)	0.02 (0.06)	0.02 (0.01)	0.07 (0.05)
Female	−0.01 (0.01)	0.003 (0.02)	0.004 (0.01)	−0.01 (0.04)	−0.004 (0.01)	0.03 (0.03)
Apprenticeship	−0.03 (0.02)	0.07 (0.06)	−0.04* (0.02)	−0.07 (0.09)	−0.01 (0.01)	0.01 (0.08)
College	−0.06*** (0.02)	0.09 (0.06)	−0.05** (0.02)	−0.04 (0.09)	−0.01 (0.01)	0.06 (0.08)
Income	−0.0002 (0.001)	0.002 (0.004)	0.0003 (0.001)	−0.003 (0.01)	0.001 (0.001)	0.002 (0.005)
Slight Interest	−0.02 (0.03)	0.002 (0.09)	0.03 (0.03)	0.23* (0.14)	−0.02 (0.02)	−0.07 (0.12)
Moderate Interest	−0.02 (0.03)	0.09 (0.08)	0.001 (0.03)	0.18 (0.12)	−0.02 (0.02)	−0.10 (0.10)
Interest	−0.02 (0.03)	0.08 (0.08)	0.02 (0.03)	0.21* (0.12)	−0.02 (0.02)	−0.06 (0.10)
High Interest	−0.01 (0.03)	0.06 (0.08)	0.0003 (0.03)	0.19 (0.12)	−0.02 (0.02)	−0.08 (0.10)
SPD	−0.05*** (0.01)	−0.10*** (0.04)	−0.01 (0.01)	−0.06 (0.05)	−0.005 (0.01)	0.05 (0.05)
FDP	0.02 (0.02)	0.001 (0.05)	0.03* (0.02)	0.05 (0.07)	−0.01 (0.01)	0.06 (0.06)
Grüne	−0.06*** (0.02)	−0.09* (0.05)	−0.05*** (0.02)	−0.16** (0.07)	−0.02* (0.01)	−0.13** (0.06)
Linke	−0.08*** (0.01)	−0.10** (0.04)	−0.04*** (0.01)	−0.10* (0.06)	−0.01 (0.01)	−0.002 (0.05)
AfD	0.08*** (0.01)	0.15*** (0.04)	0.04*** (0.01)	0.08 (0.06)	0.05*** (0.01)	0.10* (0.05)
Other	−0.02 (0.02)	−0.10* (0.05)	0.005 (0.02)	−0.16* (0.08)	0.03** (0.01)	−0.02 (0.07)
Constant	5.76*** (0.03)	5.52*** (0.10)	5.68*** (0.04)	5.95*** (0.15)	5.63*** (0.02)	5.97*** (0.13)
Observations	970	970	892	892	872	872
R^2	0.17	0.06	0.07	0.04	0.08	0.03
Adjusted R^2	0.16	0.04	0.05	0.02	0.07	0.01

*** $p < 0.01$, ** $p < 0.05$ and * $p < 0.1$. The table displays OLS estimates for six different specifications of the media slant score.

Generally, this is a clear hint that across individuals with different party identification, the selected sample of news, containing only few strictly partisan sources and mostly mainstream outlets, is consumed by everyone in comparable intensity. Once more, this challenges the perception of echo chambers in which people almost exclusively consume confirming opinions and news whilst abandoning traditional mainstream media.

As a further note on the control variables, it is surprising that age has no significant effect in 5 of the 6 specifications. In Guess (2018) age was highly significant as predictor. However, his data was collected in 2015 and focused on the US. This allows for two narratives that could explain these results. First, the penetration of online news throughout age groups could have been increased over time in 2016 and 2017 in which the US and Germany likewise experienced polarizing election campaigns. Or second, the German media landscape could be more consolidated in terms of similar content for older and younger individuals. All age groups seemingly consume similar content in Germany, whilst in the US that is not the case. This narrative would be consistent with the fact that voting liberal or conservative is much more determined by age in America than in Germany.

7 Further Evidence

Having analyzed the overlap and determinants of media diets across individuals, I turn now to the outlet level and examine news with regard to the composition of their readership. Moreover, I test a common conjecture that associates social networks with the emergence of filter bubbles and analyze a subset of observations triggered by social media.

7.1 Readers Composition

Whilst the prior analysis has shed light on the political bias of individuals media diets, it remained blurry which particular outlets can be regarded as mainstream and which are partisan with a homogeneous readership. I focused on the sample of selected news (see table 2) and analyzed the 46 outlets with regard to party identification of their readership.

The results displayed in figure 9 show the relative duration that users with a particular voting decision on election day have consumed an outlet relative to the total duration that outlet was consumed. To put these findings into perspective of total news traffic, average daily views of a domain are displayed on the right. Some findings shall be briefly pointed out.[9]

Comparing figure 9 with the political slant of outlets shows symptomatic correlations. Outlets with strong liberal ideological slant, such as *Nachdenkseiten*, *Politico*, *Neues Deutschland* and *Junge Welt* are mostly read by voters of the party *Die Linke*. On the contrary, *Sezession*, *PI News*, *Junge Freiheit*, *Compact Magazin* and *Economist* with alignment scores on the far right are read by *AfD* voters almost exclusively. Contrasting these insights with the fact that especially *Linke* and *AfD* voters exhibited ideologically aligned media diets, these additional findings enrich the evidence on what outlets are driving the divide in media consumption across political camps. However, some of these outlets are visited less than ten times a day on average, whilst the most frequently visited among the partisan domains is *PI News* with 41 average views a day. Compared to the large mainstream outlets with more than 100 up to 818 visits a day, the overall share of partisan traffic is very low.

It has to be noted that the graph does not constitute a representation of an outlets actual readership. Sample sizes can get very small for some outlets and thus introduce some bias with regard to other correlates such as voting decision. This might explain very pronounced patterns such as for *Sezession* or *Politico*, which would potentially be more moderate if measured based on a representative sample for an outlet's readership. However, the plot adds to the interpretation of earlier results and shows which outlets drive, for example, the "bumps" of *AfD* voters to the right of the distribution of media diets (see figure 5).

7.2 Social Media Analysis

Recalling that the emergence of filter bubbles is in particular associated with social media (see Bakshy et al., 2015; Barberá, 2015, and others), I contrast my findings by analyzing a subsample of news triggered through Facebook and other social networks. Recent critique has highlighted that people do not necessarily have to select themselves content they like, but through demand-driven algorithms which filter the large quantities of content and present users what matches their interests, people can involuntarily consume biased media diets.

The analysis will be conducted for the platforms *Facebook*, *Twitter* and *Reddit*. Nowadays,

[9] A list with more summary statistics and the political alignment scores of all 46 outlets is attached in the appendix.

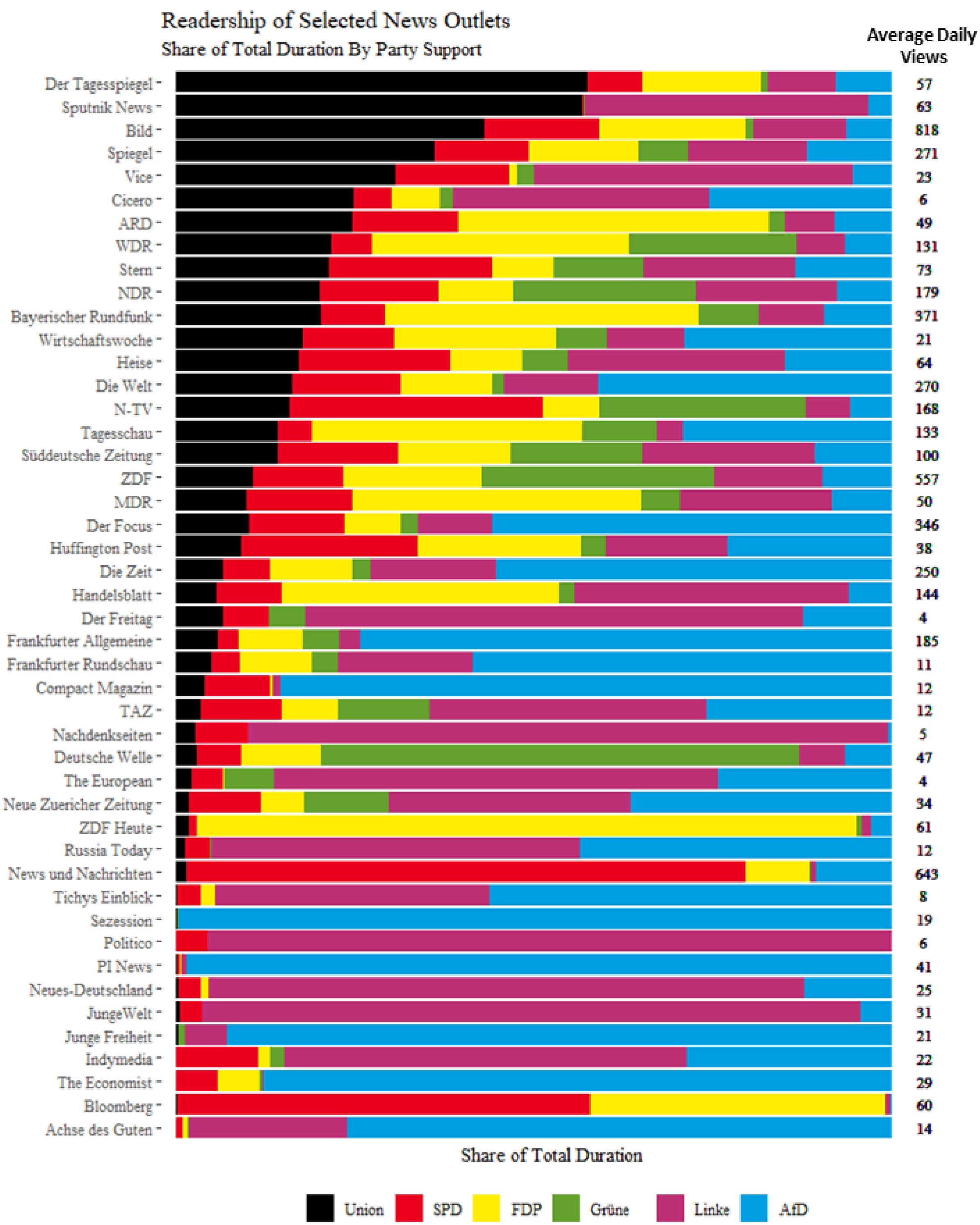


Figure 9: Outlet Readership by Party Support

the majority of people is engaged on at least one social media platform with Facebook being by far the most widely used. According to Smith and Anderson (2018), 68% of Americans use Facebook daily while Statista (2019) reports that 84% of Germans use Facebook regularly. However, consuming news on social media is less frequent. Mitchel et al. (2018) found that 26% of German adults consume news on social media of which 64% mention Facebook and 8% Twitter as primary sources. Reddit is not mentioned in this study but 4% of Germans are subscribed on Reddit and the website has an image of a more educated user base, which potentially shares different kind of content compared to Facebook and Twitter. For this reason, I decided to include all three platforms in the analysis to get a broader picture.

Subsetting of the visited domains harnesses the sequential structure of the Pulse data. I labelled every website visit as "social media triggered" when it was accessed right after a user was on Facebook, Reddit or Twitter, not exceeding a 10 seconds window after a social media website is observed. Subsequently, I subset all remaining observations by the same 2574 news domains classified for the full sample. I end up with a sample of 1245 outlets that were at least once accessed through social media. The overall share of social media induced news traffic is 3.45% compared too all news traffic in the data.

I conducted the same analysis as earlier for this restricted sample of social media triggered domains, constructed media diets as well compared them across the political spectrum. Instead of again reporting all unweighted an weighted scores, I focused on the unweighted scores for all outlets and only report the weighted scores for all other samples.

In contrast to the hypothesis of social media as facilitator of filter bubbles and news consumption in ideological echo chambers, the results almost resemble the prior analysis on the full data. Again, the vast majority of news traffic is remarkably moderate. The only caveat to an unambiguous conclusion that filter bubbles do not exist, is due to the small share of individuals who voted for *AfD*, *Linke* and *Grüne* and consume outlets on the far right and left of the political spectrum. Notwithstanding, the Facebook data even show some consumption of news to the left of the spectrum by *AfD* voters which was not evident earlier. Hence, the evidence rather supports a moderating than polarizing effect of social media with regard to overall news consumption.

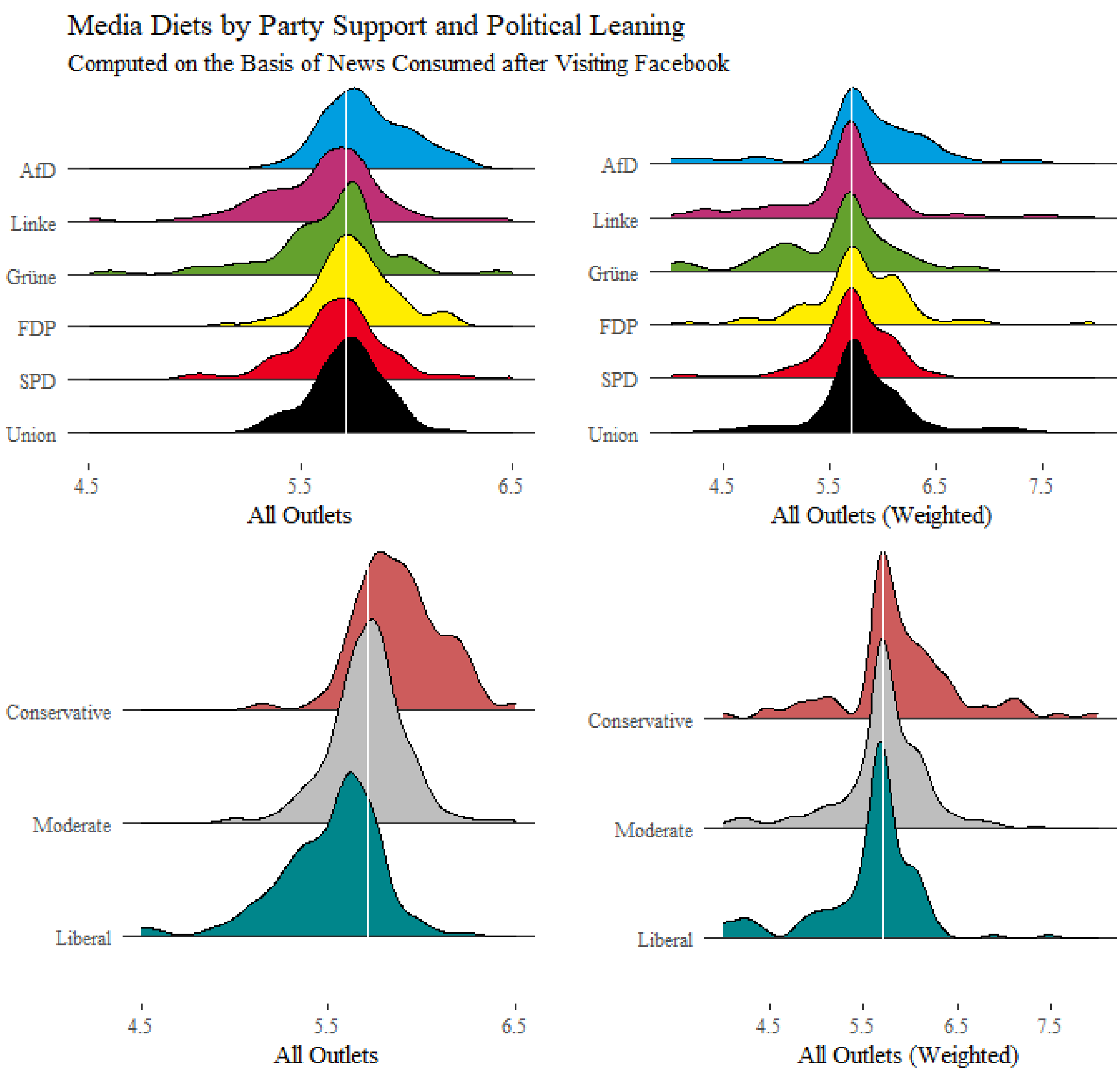


Figure 10: Social Media Triggered Media Diets

8 Limitations and Discussion

8.1 Limitations of the Research Design

Before concluding, some limitations have to be pointed out to put the overall findings into perspective. First, the study at hand used domain level data instead of full URLs. This limits the scope to which it can be controlled for the content people consumed on a website. This is less of a problem for websites such as *freitag.de* which almost exclusively focuses of politics. For other websites, like *bild.de*, this is more problematic since the domain offers not only political news, but sports, tabloid journalism and browser games. Considering all consumption of this website as news consumption, as done in this study, is unrealistic. However, a version of the Pulse data including full URLs exists and allows for a fine grained analysis that will be considered for future work.

Another issue that limits this study is the focus on data from desktop machines. As mentioned earlier, the Pew Research Center reports that 43% of Americans often read their news on mobile devices (Bialik and Matsa, 2017). Thus it is reasonable that a significant share of overall news traffic is not accounted for in the data at hand. This, however, would only bias the findings if news consumption is more slanted on mobile devices than on desktop machines. This is possible, since news might often be spread by politically aligned social networks on mobile messenger apps. But to reliably test this conjecture, more research is required.

Lastly, three more minor issues are addressed. First, for some local outlets sample sizes of readers are becoming very small. If living on the countryside is correlated with a particular political opinion, alignment scores of local news are likely biased through geographic instead of political proximity between an outlet and its readership. Second, the study does not consider news consumed directly on Facebook or Twitter. This reduces the generalizability of results on *all* news people consume. However, other studies have focused on social media and came to similar results as discussed in previous sections. Third, as highlighted earlier, this study is not designed to make causal claims on the determinants or repercussions of people's media diets. The main focus is on descriptive evidence that remained sparse in the German context. However, considering the vast amount of data and the representativeness of the sample, the results can be seen as substantial contribution illuminating who consumes what kind of news in the context of the German market of online media.

8.2 Reflecting on the Results

The study at hand has presented a large body of descriptive insights on online news consumption in Germany. The main finding is the remarkable overlap in media diets across individuals with diverging political opinion as well as differences in party support on election day. Most of the people consume classic mainstream sources and exhibit media diets centered around the median of the population. No specification of analysis has yielded convincing evidence that supports the perception of the internet as largely segregated environment of echo chambers.

Nevertheless, some individuals consume very partisan news while this pattern is more pronounced for conservatives than for liberals. Most compelling, the likelihood of people to cocoon themselves in filter bubbles does not correlate much with party identification. Rather do the results support a view that small groups of partisan individuals on the far left and right of the political spectrum tend to consume biased news. Overlap in media diets between these two poles is rather low.

Analyzing the particular readership of news outlets supports the picture that only few outlets are clearly biased. Those that show alignment scores on the far right and left are consumed by very few partisans, in particular by voters of the *AfD* and *Linke*. Finally, the overall conclusion is robust to a variant of the analysis, focusing on news consumed after the visit of social media platforms. This challenges the view that social media is a driver behind selective exposure to news.

As final note, new and innovative forms of media enter the market regularly and consumers are increasingly flexible in their habits of media consumption. All this makes persistent research difficult and poses huge challenges to future study designs. Nonetheless, considering the importance of media for democratic discourse and stability, it remains highly important to constantly improve the understanding of the repercussions of digitized media environments on people's opinion formation and eventually political change.

References

Abramowitz, A. and Saunders, K. (2005). Why Can't We All Just Get Along? The Reality of a Polarized America. *The Forum*, 3(2).

Abramowitz, A. I. and Saunders, K. L. (2008). Is Polarization a Myth? *The Journal of Politics*, 70(2):542–555.

Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science (New York, N.Y.)*, 348(6239):1130–1132.

Barberá, P. (2015). How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S. In *American Political Science Association - Annual Meeting*.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science*, 26(10):1531–1542.

Best for Planning (2017). Berichtsband b4p 2017. Technical report, Gesellschaft für integrierte Kommunikationforschung.

Bialik, K. and Matsa, E. K. (2017). Key Trends in Social and Digital News Media. Technical report, Pew Research Center.

Boxell, L., Gentzkow, M., and Shapiro, J. M. (2017). Greater Internet Use is not Associated With Faster Growth in Political Polarization Among US Demographic Groups. *Proceedings of the National Academy of Sciences of the United States of America*, 114(40):10612–10617.

Cage, J. and Goldhammer, A. (2016). *Saving the Media: Capitalism, Crowdfunding, and Democracy*. Harvard University Press.

Clemons, T. E. and Bradley, E. L. (2000). A Nonparametric Measure of the Overlapping Coefficient. *Computational Statistics & Data Analysis*, 34(1):51–61.

Dahl, R. A. (2008). *On Democracy*. Yale University Press, New Haven & London.

Dallmann, A., Lemmerich, F., Zoller, D., and Hotho, A. (2015). Media Bias in German Online Newspapers. In *26th ACM Conference on Hypertext & Social Media*, pages 133–137, New York, New York, USA. ACM Press.

Dellavigna, S. and Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting. *Quarterly Journal of Economics*, 122(3):1187–1234.

Dubois, E. and Blank, G. (2018). The Echo Chamber is Overstated: The Moderating Effect of Political Interest and Diverse Media. *Information, Communication & Society*, 21(5):729–745.

Eady, G., Nagler, J., Guess, A., Zilinsky, J., and Tucker, J. A. (2019). How Many People Live in Political Bubbles on Social Media? Evidence From Linked Survey and Twitter Data. *SAGE Open*, 9(1):215824401983270.

Fiorina, M. P., Abrams, S. A., and Pope, J. C. (2008). Polarization in the American Public: Misconceptions and Misreadings. *The Journal of Politics*, 70(2):556–560.

Fiorina, M. P., Abrams, S. J., and Pope, J. (2011). *Culture war? The Myth of a Polarized America*. Longman.

Flaxman, S., Goel, S., Rao, J. M., Blei, D., Budak, C., Dumais, S., Gelman, A., Goldstein, D., Salganik, M., Wu, T., and Zervas, G. (2016). Filter Bubbles, Echo Chambers, and Online News Consumption. *Public Opinion Quarterly*, 80(Special Issue):298–320.

Füchtjohann, J. (2018). Demokratie muss rumpeln und stottern.

Gastwirth, J. L. (1975). Statistical Measures of Earnings Differentials. *The American Statistician*, 29(1):32.

Gentzkow, M. and Shapiro, J. M. (2010). What Drives Media Slant? Evidence from U.S. Daily Newspapers. *Econometrica*, 78(1):35–71.

Groseclose, T. and Milyo, J. (2005). A Measure of Media Bias. *The Quarterly Journal of Economics*, 120(4):1191–1237.

Guess, A. M. (2015). Measure for Measure: An Experimental Test of Online Political Media Exposure. *Political Analysis*, 23(1):59–75.

Guess, A. M. (2018). (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets.

Habermas, J. (1962). *The Structural Transformation of the Public Sphere: An Inquiry Into a Category of Bourgeois Society*.

Habermas, J. (1984). *The Theory of Cummunicative Action*. Beacon Press.

Hare, C., Armstrong, D. A., Bakker, R., Carroll, R., and Poole, K. T. (2015). Using Bayesian Aldrich-McKelvey Scaling to Study Citizens' Ideological Preferences and Perceptions. *American Journal of Political Science*, 59(3):759–774.

Inman, H. F. and Bradley, E. L. (1989). The Overlapping Coefficient as a Measure of Agreement Between Probability Distributions and Point Estimation of the Overlap of

Two Normal Densities. *Communications in Statistics - Theory and Methods*, 18(10):3851–3874.

Klapper, J. T. (1960). *The effects of mass communication: An Analysis of Research on the Effectiveness and Limitations of Mass Media in Influencing the Opinions, Values, and Behavior of their Audiences.* Free Press.

Lelkes, Y. (2016). Mass Polarization: Manifestations and Measurements. *Public Opinion Quarterly*, 80(S1):392–410.

Levendusky, M. S. and Pope, J. C. (2011). Red states vs. blue states: Going beyond the mean. *Public Opinion Quarterly*, 75(2):227–248.

Lobe, A. (2018). Was passiert, wenn die Filterblase platzt?

Martin, G. J., Yurukoglu, A., Clark, T., Crawford, G., Enikopolov, R., Gentzkow, M., Golub, B., Hinnosaar, M., Kawai, K., Lee, R., Lim, C., Oyer, P., Pakes, A., Shapiro, J., Sinkinson, M., and Sood, G. (2017). Bias in Cable News: Persuasion and Polarization. *American Economic Review*, 107(9):2565–2599.

Messing, S. and Westwood, S. J. (2014). Selective Exposure in the Age of Social Media. *Communication Research*, 41(8):1042–1063.

Mitchel, A., Matsa, K. E., Shearer, E., Johnson, C., Simmons, K., Silver, L., Walter, M., and Taylor, K. (2018). In Western Europe, Public Attitudes Toward News Media More Divided by Populist Views Than Left-Right Ideology. Technical report, Pew Research Center.

Munzert, S., Barberá, P., Guess, A., and Yang, J. (2017). Media Exposure and Opinion Formation in an Age of Information Overload. In *American Political Science Association - Annual Meeting*, page 30, San Francisco.

Mutz, D. and Martin, P. (2001). Facilitating Communication Across Lines of Political Difference: The Role of Mass Media. *American Political Science Review*, 95(1):97–114.

Nordenstreng, K. (2000). Media and Democracy: What is Really Required? In van Cuilenburg, J. and van der Wurff, R., editors, *Media & Open Societies: Cultural, Economic and Policy Foundations for Media Openness and Diversity in East and West*, pages 29–47. Het Spinhuis, Amsterdam.

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. The Penguin Press, New York, 2 edition.

Pew Research Center (2018). Trends and Facts on Online News | State of the News Media | Pew Research Center.

Prior, M. (2007). *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge University Press.

Prior, M. (2013). Media and Political Polarization. *Annual Review of Political Science*, 16(1):101–127.

Reich, R. B. (1988). *The Power of Public Ideas*. Ballinger.

Rivers, D. (2006). Sample Maching: Representative Sampling from Internet Panels.

Schmid, F. and Schmidt, A. (2006). Nonparametric Estimation of the Coefficient of Overlapping - TTheory and Empirical Application. *Computational Statistics & Data Analysis*, 50(6):1583–1596.

Smith, A. and Anderson, M. (2018). Social Media Use in 2018. Technical report, Pew Research Center.

Statista (2019). Which Social Networks Do You Use Regularly?

Steveson, A. (2018). Facebook Admits It Was Used to Incite Violence in Myanmar.

Stroud, N. J. (2008). Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure. *Political Behavior*, 30(3):341–366.

Sunstein, C. R. (2017). *#Republic: Divided Democracy in the Age of Social Media*.

Tufekci, Z. (2018). YouTube, the Great Radicalizer.

Vaccari, C., Valeriani, A., Barberá, P., Jost, J. T., Nagler, J., and 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).

Yanagizawa-Drott, D. (2014). Propaganda and Conflict: Evidence from the Rwandan Genocide. *The Quarterly Journal of Economics*, 129(4):1947–1994.

A Dictionaries

Customized News Dictionary

abendblatt, abendzeitung-muenchen, achgut, augburger-allgemeine, badische-zeitung, bento, berliner-zeitung, bild, braunschweiger-zeitung, businessinsider, bz-berlin, compact-online, daserste, deutschlandfunk, deutschlandfunkkultur, express, faz, focus, fr, freiewelt, funkemedien, handelsblatt, hessenschau, heute, hna, huffingtonpost, infranken, junge-freiheit, jungle, kenfm, ksta, manager-magazin, mdr, mediengruppe-thueringen, merkur, mmnews, mopo, morgenpost, mz-web, n24, national-zeitung, ndr, news, news-und-nachrichten, nordbayern, noz, n-tv, nw, nzz, ovb24, ovb-online, pi-news, pnp, punto-informatico, rnd-news, rp-online, sezession, shz, spiegel, stern, stuttgarter-nachrichten, sueddeutsche, tag24, tagesschau, tagesspiegel, taz, tichyseinblick, tv-mittelrhein, tz, unzensuriert, up-day, vice, wdr, welt, wiwo, zeit

Customized List of Regex Expressions

journal, journalisten, magazin, news, rundschau, zeitung

B Media Slant

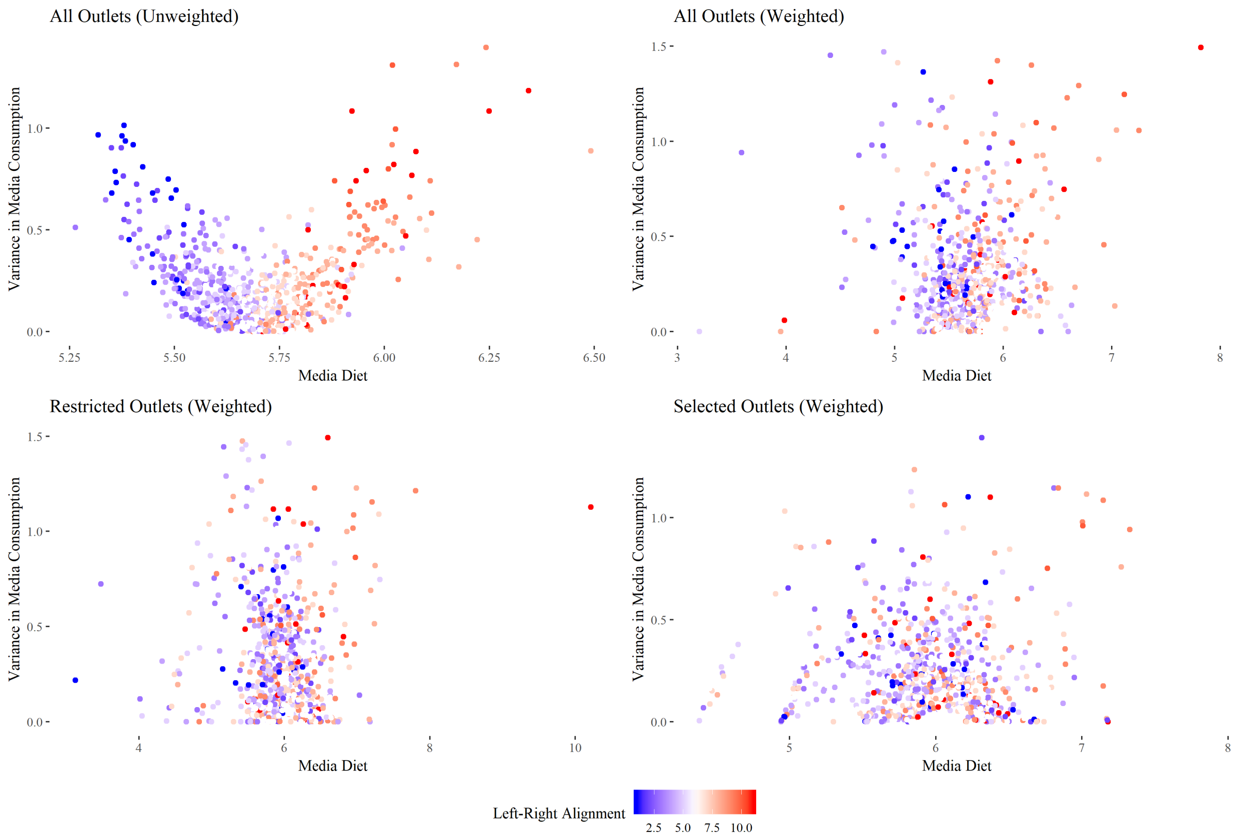
Table 6: Media Slant of Selected Outlets

Domain	Slant	Slant Weighted	Duration (hrs)	Views (1000)	Users	Traffic (%)
Bild	5.79	6.38	2155.18	128.702	685	22.10
ZDF	5.49	4.94	1107.80	74.147	415	11.36
Der Focus	5.67	5.85	956.45	53.721	800	9.81
Die Welt	5.71	6.49	870.74	39.088	713	8.93
Spiegel	5.58	5.63	643.29	38.409	664	6.60
News und Nachrichten	5.69	4.38	557.17	96.211	39	5.71
Die Zeit	5.41	6.74	487.14	35.164	435	4.99
Frankfurter Allgemeine	5.57	7.18	421.79	26.990	406	4.32
N-TV	5.61	5.53	393.36	24.493	380	4.03
Tagesschau	5.51	6.25	267.88	17.446	243	2.75
WDR	5.52	6.29	189.10	14.145	271	1.94
Der Tagesspiegel	5.41	6.35	183.89	6.533	303	1.89
Süddeutsche Zeitung	5.46	5.30	143.08	11.046	446	1.47
ZDF Heute	5.67	8.06	143.02	5.356	117	1.47
Stern	5.69	5.81	139.01	10.020	513	1.43
NDR	5.50	5.52	119.80	12.043	358	1.23
Handelsblatt	5.76	6.65	118.15	9.664	288	1.21
PI News	6.95	8.37	98.78	4.286	39	1.01
Huffington Post	5.61	5.83	93.48	4.954	341	0.96
Bayerischer Rundfunk	5.65	6.21	92.65	17.698	237	0.95
MDR	5.45	6.69	74.94	5.964	250	0.77
Sputnik News	6.09	7.36	66.21	7.872	50	0.68
ARD	5.42	7.17	51.90	5.969	244	0.53
Heise	5.74	5.20	46.29	6.228	425	0.47
JungeWelt	5.33	5.35	34.38	1.956	25	0.35
Junge Freiheit	7.23	7.94	32.75	1.954	41	0.34
Bloomberg	6.21	5.41	32.67	3.157	29	0.33
Vice	5.26	4.75	28.25	1.731	158	0.29
Achse des Guten	5.78	7.05	23.43	0.880	28	0.24
Russia Today	5.96	7.60	22.81	1.069	59	0.23
The Economist	5.89	7.29	20.89	1.979	19	0.21
Frankfurter Rundschau	5.58	7.12	18.30	1.085	112	0.19
TAZ	5.36	5.51	15.22	0.959	133	0.16
Wirtschaftswoche	5.76	6.35	14.59	1.518	160	0.15
Neues-Deutschland	5.15	4.09	13.89	0.795	48	0.14
Tichys Einblick	6.48	6.04	13.67	0.420	33	0.14
Deutsche Welle	5.44	7.91	11.68	0.874	55	0.12
Neue Zuericher Zeitung	6.04	5.73	10.88	0.755	75	0.11
Sezession	7.40	7.28	9.85	1.573	5	0.10
Compact Magazin	6.86	7.83	8.18	0.260	22	0.08
Cicero	5.97	5.89	5.57	0.248	40	0.06
Nachdenkseiten	5.06	6.43	4.66	0.206	19	0.05
The European	6.05	5.40	4.02	0.121	22	0.04
Der Freitag	5.00	4.02	3.58	0.159	32	0.04
Indymedia	6.13	5.95	2.88	0.192	24	0.03
Politico	5.33	3.52	0.51	0.050	6	0.01

Table displays alignment scores of selected outlets. Low scores indicate liberal and high scores conservative bias. Traffic is measured as percentage of total duration outlets were consumed.

C Additional Results

Scatterplot of Mean and Variance of Individual Media Diets



Mean vs. Variance in Individuals' Media Diets Colored By Left-Right Alignment

Regressions on Political Leaning Instead of Party Support

Regression Analysis for Categorical Political Leaning

	All Outlets Raw	All Outlets Weighted	Restricted Outlets Raw	Restricted Outlets Weighted	Selected Outlets Raw	Selected Outlets Weighted
	(1)	(2)	(3)	(4)	(5)	(6)
Age: 30-39	0.02 (0.01)	0.004 (0.04)	0.01 (0.01)	−0.09 (0.06)	0.03*** (0.01)	−0.0004 (0.05)
Age: 40-49	0.03** (0.01)	0.0002 (0.04)	0.003 (0.01)	−0.06 (0.06)	0.02** (0.01)	0.05 (0.05)
Age: 50-59	0.03*** (0.01)	0.03 (0.04)	0.01 (0.01)	−0.04 (0.06)	0.02*** (0.01)	−0.01 (0.05)
Age: >60	0.02 (0.01)	0.03 (0.04)	0.01 (0.01)	0.0000 (0.06)	0.01 (0.01)	0.03 (0.05)
Female	−0.005 (0.01)	0.02 (0.02)	0.004 (0.01)	0.01 (0.03)	−0.004 (0.01)	0.05* (0.03)
Apprenticeship	−0.01 (0.02)	0.07 (0.05)	−0.02 (0.02)	−0.0005 (0.08)	−0.01 (0.01)	−0.003 (0.07)
College	−0.04** (0.02)	0.07 (0.06)	−0.04** (0.02)	0.01 (0.08)	−0.01 (0.01)	0.04 (0.07)
Income	−0.001 (0.001)	0.001 (0.004)	0.0000 (0.001)	0.003 (0.01)	0.0005 (0.001)	0.005 (0.005)
Slight Interest	−0.004 (0.02)	−0.09 (0.08)	0.03 (0.03)	0.26** (0.13)	−0.03 (0.02)	0.02 (0.11)
Moderate Interest	−0.02 (0.02)	−0.01 (0.07)	−0.02 (0.03)	0.25** (0.11)	−0.03* (0.02)	0.04 (0.10)
Interest	−0.01 (0.02)	−0.01 (0.07)	0.01 (0.03)	0.31*** (0.11)	−0.03 (0.02)	0.08 (0.10)
High Interest	−0.01 (0.02)	−0.03 (0.07)	−0.01 (0.03)	0.29** (0.11)	−0.03* (0.02)	0.07 (0.10)
Liberal	−0.16*** (0.01)	−0.16*** (0.03)	−0.08*** (0.01)	−0.13*** (0.05)	−0.03*** (0.01)	−0.05 (0.04)
Conservative	0.16*** (0.01)	0.20*** (0.04)	0.05*** (0.01)	0.20*** (0.06)	0.02*** (0.01)	0.17*** (0.05)
Constant	5.73*** (0.03)	5.60*** (0.09)	5.69*** (0.03)	5.71*** (0.14)	5.65*** (0.02)	5.83*** (0.12)
Observations	1,131	1,131	1,024	1,024	998	998
R^2	0.35	0.06	0.09	0.04	0.05	0.02
Adjusted R^2	0.34	0.05	0.08	0.02	0.04	0.01

*** $p < 0.01$, ** $p < 0.05$ and * $p < 0.1$. The table displays OLS estimates for six different specifications of the media slant score.

Statement of Authorship

I hereby confirm and certify that this master thesis is my own work. All ideas and language of others are acknowledged in the text. All references and verbatim extracts are properly quoted and all other sources of information are specifically and clearly designated. I confirm that the digital copy of the master thesis that I submitted on April 29, 2019 is identical to the printed version I submitted to the Examination Office on April 30, 2019.

Berlin, April 29, 2019 ....