P&I
POLICY & INTERNET

Check for updates

Measuring Proximity Between Newspapers and Political Parties: The Sentiment Political Compass

Fabian Falck, Julian Marstaller, Niklas Stoehr, Sören Maucher, Jeana Ren, Andreas Thalhammer, Achim Rettinger, and Rudi Studer

The proximity between newspapers and political parties is strongly subjective and difficult to measure. Yet, political tendencies of newspapers can have a significant impact on voters' opinion-forming and ought to be known by the public in a transparent and timely manner. This article introduces the Sentiment Political Compass (SPC), a data-driven framework for analyzing political bias of newspapers toward political parties. Using the SPC, newspapers are embedded in a two-dimensional space (left-leaning vs. right-leaning, libertarian vs. autocratic). To assess the informative value of our framework, we crawled a data set consisting of 180,000 newspaper articles from twenty-five newspapers during the German Federal Elections over a time period of 18 months and extracted 740,000 political entities enriched with their contextual sentiment. We analyze this dataset on the party- and politician-level as well as considering the temporal dimension and draw insights about the relationship between newspapers and political parties. We provide the data set and our code open-source at www.politicalcompass.de to encourage the application of the SPC to other political landscapes.

KEY WORDS: political compass, democratic transparency, sentiment analysis, news media, search engine crawling, opinion mining, political attitudes, machine learning, data mining

报刊和政党之间的亲近程度是极具主观性且难以衡量的。尽管如此，报刊的政治倾向能对选民的意见形成产生显著影响，因此应该以一种透明且及时的方式被公众所熟知。本文引入"情感政治倾向测试"*(SPS)*,一项由数据驱动，用于分析报刊对政党所持的政治偏见的框架。通过使用*SPS*，报刊被置于一个二维空间（左倾与右倾，自由与专制）。为评估该框架的信息值，我们花费 *18* 个月的时间对一个数据集进行了搜索引擎爬行，该数据集由德国联邦选举期间 *25* 家报社发行的 *180000* 篇报刊文章组成，我们从中提取了 *740000* 个富含文本情感的政治实体。我们从党派和政客的层面分析该数据集，同时考量时间维度，并对报刊和政党之间的关系提炼出见解。我们在网站 *www.politicalcompass.de* 上提供了该数据集和代码开放源，以期鼓励 *SPC* 被应用于其他政治格局。

关键词： 情感政治倾向测试, 情感分析, 意见挖掘, 新闻媒体, 搜索引擎爬行, 民主, 政治偏见

La proximidad entre periódicos y partidos políticos es fuertemente subjetiva y difícil de medir. Sin embargo, las tendencias políticas de los periódicos pueden tener un impacto significativo en la

doi: 10.1002/poi3.222
© 2019 Policy Studies Organization

formación de opinión de los votantes y deben ser conocidas por el público de manera transparente y oportuna. Este artículo presenta el Sentiment Political Compass (SPC), un marco basado en datos para analizar el sesgo político de los periódicos hacia los partidos políticos. Usando el SPC, los periódicos se incrustan en un espacio bidimensional (inclinado a la izquierda versus inclinado a la derecha, libertario versus autocrático). Para evaluar el valor informativo de nuestro marco, rastreamos un conjunto de datos que constaba de 180,000 artículos periodísticos de 25 periódicos durante las elecciones federales alemanas durante un período de tiempo de 18 meses y extrajimos 740,000 entidades políticas enriquecidas con su sentimiento contextual. Analizamos este conjunto de datos a nivel de partido y nivel político, así como teniendo en cuenta la dimensión temporal, y extraemos información sobre la relación entre los periódicos y los partidos políticos. Proporcionamos el conjunto de datos y nuestro código abierto en www.politicalcompass.de para alentar la aplicación del SPC a otros paisajes políticos.

PALABRAS CLAVE: brújula política, análisis de opiniones, minería de opinión, medios de comunicación, rastreo de motores de búsqueda, democracia, sesgo político

Introduction

Political Bias of Online Newspapers

News media is known as the fourth pillar in the democratic process, taking the position of a mediator between politics and society (Beck, 2012). In assuming this powerful role, it may strongly influence public beliefs and act as an opinion maker, as evidenced by many historic election campaigns. Recently, online news media have been accused of being biased toward particular political positions and standpoints, with terms like "fake news" or "AfD bashing" assuming an inherent bias of newspapers (Bessi et al., 2015; Gaebler, 2017). However, in most cases, voters have only a vague and subjective perception of a newspaper's proximity to a political party. If voters are uninformed about the political point of view of media reporting, they may be manipulated in their democratic opinion-forming (Mutz, 1989). Therefore, we propose that qualitative efforts to classify political conviction and bias need to be augmented by more quantitative models. This study aims to answer the following research question: How can we measure and analyze the political biases of newspapers and their proximity to political parties with machine-learning methods?

The Sentiment Political Compass

One possible way to address biased media reporting and supposedly false information is through greater transparency and quantifiability. This article analyses newspapers with respect to their political orientation by making the following three contributions:

(a) We present the *Sentiment Political Compass (SPC)*, a data-driven framework that classifies the attitude of newspapers toward political parties. The approach is transparent, being based on entity sentiment analysis of thousands of

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

newspaper articles to reveal the temporal connection between political events and societal shifts in opinion. These connections are visualized in a two-dimensional space resembling a compass (left vs. right, libertarian vs. autocratic), which serves as a tool for media monitoring.

(b) We present the technical details of the system including the framework used to crawl newspaper articles, extract entities, and perform entity sentiment analysis. We provide two implementations of the entire working pipeline, one version using commercial cloud services and another being entirely open source. For this reason, the Sentiment Political Compass may be reproduced and applied to analyze the relation between the media landscape and politics in any country in the world.

(c) We demonstrate the informative value of our approach by analyzing the political media sentiment in Germany during the year of the German Federal Parliament and subsequent coalition negotiations in 2017/2018. Our final data set contains over one hundred eighty thousand newspaper articles from eighteen newspapers and seven party-affiliated newspapers, with around seven hundred forty thousand matched entity sentiments, that is, sentiment scores matched with an entity subject in the text. Our analysis yields new insights into the distribution of media coverage per political party, and overall media sentiment.

Our data retrieval and analysis revolves around the construct *newspaper*, which is assumed to be the object of observation to measure political bias. However, note that newspaper articles from a single publishing house are typically written by numerous authors and editors who might have and express differing views in their work. Theoretically, we argue that the existence of the group (i.e. the newspaper) influences the individuals (authors) to write with a cohesive political voice. Many examples of the influence of groups on their component individuals can be found in the psychological literature, for example, on decision making for the sake of conformity within the group ("group-thinking") (Janis, 2008), the transformation of individual positions toward a more extreme (group) position ("group shift") (Silverthorne, 1971), or loss of self-awareness and de-individuation (Diener, 1980; LumenLearning, 2019). This is why we analyze the political bias at the level of the group (i.e. the newspaper).

The newspaper articles used for the analysis include both editorials and news articles from central agencies, such as the German Press Agency (DPA). We argue that both article types will reflect an underlying political proximity (i.e. political slant or bias) of a group. In an editorial, the proximity emerges directly through expressed sentiments, while externally purchased news articles represent the coverage bias of newspapers selecting articles from the daily available press corpus.

The general working pipeline of the Sentiment Political Compass, illustrated in Figure 1, involves three high-level steps: (i) *Article crawling*: extraction of newspaper articles; (ii) *Entity extraction and sentiment analysis*: identification of political

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

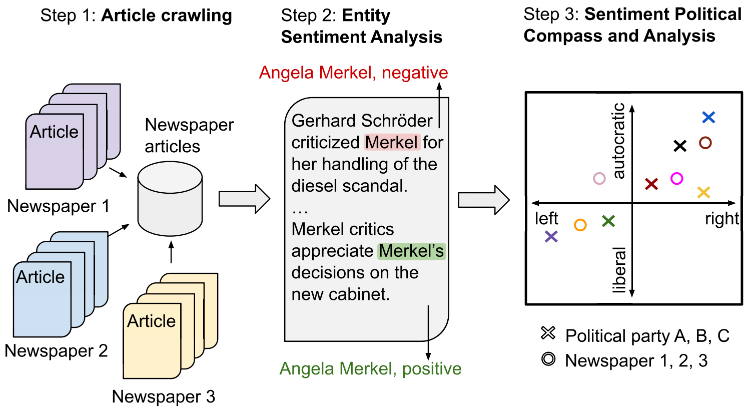

Figure 1. Schematic Illustration of the High-Level Steps to Perform the Entity Sentiment Analysis and Its Subsequent Analysis, Concluding in the Sentiment Political Compass.

entities in newspaper articles and analysis of their contextual sentiment; (iii) *Sentiment Political Compass*: computation of the political position of the newspapers and subsequent analysis. Even though we present a generic framework that is applicable to any country and time frame, we refer specifically to our crawled data set throughout the paper for ease of explanation.

This article is structured as follows: First, we introduce related work and shed light on the technical details of the framework. This involves a description of how to crawl the newspaper articles, retrieve entities, and perform entity sentiment analysis. These steps form the basis of our data pipeline. Then, we present extensive analytical results, concluding in the definition and construction of the Sentiment Political Compass.

Related Work

Classification of Political Conviction

Dimensions of Political Opinion. In this section, we present models and fundamental research in political science that form the basis of the Sentiment Political Compass. The terms left-wing and right-wing date back to the French revolutionary era (Lester, 1994); however, the first attempts to quantify political conviction originate from the 1940s, when Leonard W. Ferguson and Hans Eysenck designed models for a factor analysis of political values (Eysenck, 1957; Ferguson, 1941). Nolan (1969) created a chart diagram contrasting four major political convictions—conservative and liberal, authoritarian and libertarian—in a two-dimensional space. According to the *Nolan Chart*, a differentiating criterion is a party's view toward economic and personal freedom. Numerous successors and higher dimensional variants of the Nolan Chart exist, for example, the models introduced by Meltzer and Christie (1970) and Bryson and McDill (1968). Fritz (1987) first developed a model that introduces political parties into the Nolan Chart: "The world's smallest political

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

quiz" assigns a position in the plot according to a test taker's answers to 10 questions divided into the economic and the personal (Fritz, 1987). Greenberg and Jonas (2003) constructively discussed "psychological motives and political orientation" and laid the foundation of the *political compass*, a two-dimensional model introduced by the organization politicalcompass.org (politicalcompass.org, 2001). The political compass constitutes a restructured version of the Nolan Chart by shifting back its interpretation to the standard leftright axis and a vertical axis representing ideological rigidity. Even though the political compass still tries to capture the political landscape in a two-dimensional space only, it combines long-established terminology with a more comprehensive classification that allows for quantitative analysis.

Quantifying Political Opinion. In order to translate political views into a two-dimensional space, the political compass builds upon Fritz's rule-based quiz by scoring the answers to sixty-one questions on the scale "strongly agree," "agree," "disagree," and "strongly disagree." In a similar fashion, the German Federal Agency for Civic Education offers a question-and-answer website to let voters compare their political conviction on specific topics with those of political parties (Federal Centre for Political Education of Germany, 2002). The so-called *Wahl-o-Mat* uses a scoring framework that resembles "The world's smallest political quiz".

Politicalcompass.org provided a finalized model of the German party landscape in the 2017 election year. As regards the Sentiment Political Compass, our work builds on this existing model, which classifies political parties by enriching it with the political orientation of newspaper articles toward political parties based on sentiment information. Politicalcompass.org assess the leftright dimension based particularly on economic measures, as reflected in their survey questions: thus, "far right" implies neo-liberalism, not necessarily nationalism. In their 2017 evaluation of the German federal election, politicalcompass.org placed the AfD less "economically" right compared with FDP and CDU, while in public discourse, the AfD was typically classified as a right-wing party. This ambiguity results from the existing difference in the terminology. The public discourse is often centered around the parties' political stance toward current societal issues like the refugee crisis or the relationship between Europe and the nation state. In this work, we strictly follow the economic definition of politicalcompass.org when developing our final model, as we utilize their evaluation of political parties.

Sentiment Analysis Regarding Political Orientation

Technical Aspects of Sentiment Analysis. Having discussed scientific approaches to measuring political conviction, we now aim at gaining a more detailed view on sentiment analysis in news media. *Sentiment Analysis* is an established term in natural language processing and is sometimes referred to as "opinion mining" in the context of newspaper analysis (Pang et al., 2008). Sentiment analysis is the procedural attempt of computationally identifying opinions expressed in written text. Pang, Lee, and Vaithyanathan (2002) studied the suitability of different

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

machine-learning approaches for sentiment analysis. The Sentiment Political Compass takes their findings into account by making use of *entity sentiment analysis*, which exploits rich semantics to analyze individual entities in a text in comparison with a text as a whole. Entity sentiment analysis has been extensively surveyed (Ravi & Ravi, 2015) and applied to political opinion mining and tendency identification (Pla & Hurtado, 2014).

Methods and Data Sources for Measuring Political Opinions. The "subjectivity" in speech toward different political opinions has been studied extensively in various contexts (Carbonell, 1979; Wilks & Bien, 1983). With the emergence of online newspapers and social media, quantitative methods have replaced these early qualitative foundations. Anstead and O'Loughlin (2014) generally address the relationship between social media and public opinion through the prism of public opinion theory. An extremely popular platform for sentiment analysis is Twitter. There exist numerous generic technical frameworks to analyze the political alignment of Twitter users (Gao & Sebastiani, 2015; Gautam & Yadav, 2014; Pla & Hurtado, 2014). Gautam and Yadav (2014) use Twitter as a test bed for machine learning algorithms in combination with WordNet's semantic orientation. Other work discusses how sentiment classifications from a few data entities may be quantified (Gao & Sebastiani, 2015), or describes a machine-learning pipeline for identifying political tendencies in tweets (Pla & Hurtado, 2014). Some researchers have debated the potential of Twitter analysis to replace traditional polls (Chung & Mustafaraj, 2011; Gayo-Avello, 2013). Other work has exploited Facebook posts for prediction and analysis of political conviction (Caton, Hall, & Weinhardt, 2015; Neri, Aliprandi, Capeci, Cuadros, & By, 2012). For instance, Neri et al. (2012) compare Facebook posts and their sentiments about newscasts from two broadcasting companies, and Caton et al. (2015) quantitatively investigate how politicians use Facebook. Indicating the variety of potential media sources, Holtzman, Schott, Jones, Balota, and Yarkoni (2011) explore media bias in television transcripts using semantic analysis tools. In contrast to the related work we discuss above, the data sources and media we exploit are freely accessible online newspapers.

Applications to Political Events. A rich corpus of work on prediction and analysis focuses on specific events with a large political scope. Since our work examines the political landscape in Germany, we particularly highlight a few papers concerned with the German Federal Elections in 2009 (Jungherr, Jürgens, & Schön, 2011; Tumasjan, Sprenger, Sandner, & Welpe, 2010), in 2013 (Jungherr, Schön, & Jürgens, 2013; Kaczmirek et al., 2014) and in 2017 (Stier et al., 2018). The working paper by Stier et al. (2018) like the present study, analyzes the German Federal Elections of 2017. However, the authors neither examine newspapers, nor perform sentiment analysis. In fact, most of the work listed above analyzes political opinions on social media platforms instead of online newspapers. We believe that one reason for this bias in research is the ease of data availability through standardized APIs, particularly on Twitter (Twitter, Inc., 2019), which is why a detailed discussion of technical frameworks to gather

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

other politically relevant and influencing information, such as newspaper articles, is opening up potential future branches of research.

Besides the German Federal Elections, various other major political events have been closely monitored using quantitative approaches, including the US Presidential Elections (2008) (Malouf & Mullen, 2008), the Arab Spring in Egypt (2011) (Boecking, Hall, & Schneider, 2015), the General Elections in Belgium (2011) (De Fortuny, De Smedt, Martens, & Daelemans, 2012), the Presidential Ballot in France (2011) (Ceron, Curini, Iacus, & Porro, 2014), the Independence Referendum in Scotland (2014) (Wagner, 2018), the Brexit referendum (2016) (Hurlimann et al., 2016), and the constitutional referendum in Italy (2016) (Marozzo & Bessi, 2017)—to name just a few.

In the context of the 2013 Austrian parliamentary election campaign, the partisan bias of newspapers was analyzed through bias in coverage, by comparing the frequency that party press releases were included in the news articles (Haselmayer, Wagner, & Meyer, 2017). Haselmayer et al. showed that the likelihood of a newspaper covering a party press release correlated positively with the party orientation of its readership, drawn from the surveys. Proximity to a political party was therefore expressed through the increased attention of a newspaper toward that party, resulting in increased coverage. Our own study aims at disclosing coverage bias as well, but by measuring the frequency of mentions of party entities. Additionally, we believe that extracting the sentiment toward a mentioned entity is a relevant extension of the frequency metric to measure the coverage bias contributing to political proximity, as it is unaffected by a newspaper's potential intention to cover more content of a certain party only to refute their views or even to traduce them.

To the best of our knowledge, there exists no prior work assessing the relationship between newspaper media and political parties using entity sentiment analysis on a large amount of crawled newspaper articles.

Technical Framework

This section describes the data pipeline to generate a corpus of semantically labeled entities on which all subsequent analysis will be based on. The high-level steps of this technical framework are as follows: First, we construct a database of political entities from different categories (such as party names or political representatives) and a second database of mass-media newspapers with diverse political orientations. Then, we search for a subset of these entities on selected newspaper domains in dynamically adjusted time periods and gain article URLs stored in a database. The entity categories and whether they are used as search terms are listed in Table 1. We also extract various meta-information from the articles (such as the article date). Subsequently, we extract the raw entities from the text and match them with our previously defined entity corpus. Finally, the sentiments of these entities based on their context within the article are computed. This last step relies on the state-of-the-art machine-learning libraries and approaches. An overview of

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1. Entity Corpus Categories

Entity Category	# of Unique Entities	Search Term
Party name	10	✓
Youth organization name of party	7	✓
Chairperson	9	✓
Deputy chairperson	31	✓
Chancellor	1	✓
Minister of the federal government	21	✓
Secretary general	6	✓
State premier	16	✓
Member of federal parliament	918	
Honorary chairperson	12	

Note that if an entity belongs to two groups (e.g., is a member of the federal parliament and a member of the government), the categorization with the least entities (here member of government) is listed. If an entity is present in two different legislative periods, it is listed only once. Any entity occurring in an article that is also in the entity corpus is enriched with a sentiment. A subset of the entity corpus is used as search terms.

the data pipeline, which will now be discussed in detail is given in Figure 2. Note that the six steps in the figure correspond to the following six subsections.

Online Newspaper Selection

We analyzed both qualitative and quantitative criteria to decide which online newspapers, political magazines, and party-related information websites should be chosen for data crawling. For a *qualitative* selection, we put particular emphasis on independent, nationwide newspapers. Although local newspapers play an important role in the German media landscape, they often discuss the content that is relevant to a locally restricted group of people only. Moreover, local newspapers

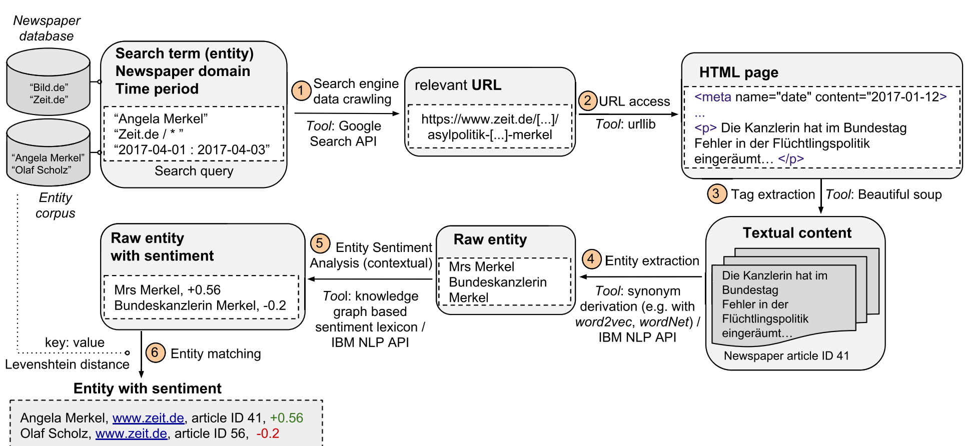

Figure 2. Schematic Illustration of the Pipeline to Obtain the Data Underlying Our Analysis.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

tend to duplicate content that is relevant to a wider audience from a sister newspaper within the same publisher, such as on topics regarding foreign affairs or economic issues (Butenschön et al., 2017; Maute, 2011). Despite of their high net reach, we exclude news streaming services, such as upday (Samsung) or Google news, which refer to articles from other sources and publish only a small fraction of their own content. As a sanity check, we take presumably highly biased party newspapers and magazines of each major political party as well as a few extremist newspapers into account, despite their smaller audience. We also considered the public opinion of the newspaper's political orientation to ensure that the complete political spectrum is homogeneously represented. Note that the consistency between the newspaper's data-inferred political orientation and the public opinion on its orientation is one major research question to be analyzed by this study, which is why we consider the (rather subjective) latter factor only for selection purposes, not as the "ground truth".

Furthermore, two *quantitative* criteria both assessing the popularity of newspapers were evaluated. First, we considered the average number of monthly *PageImpressions* within the observation period from the category online offerings restricted to image and text content in German (German audit bureau of circulation, 2018c), as measured by the German Audit Bureau of Circulation (IVW), an "independent, noncommercial and neutral auditing agency" (German audit bureau of circulation, 2018a). PageImpressions are defined as clicks on a website unequivocally assigned to the newspaper by its fully qualified domain (FQDN) or an alias/redirect and explicitly caused by user input (German audit bureau of circulation, 2018b). As a second criterion, we analyzed the net reach of online newspapers in terms of unique users (statista.com, 2018).

Entity Selection

In endeavoring to analyze a newspaper's sentiment toward a political party, we need to establish a corpus of entities that are associated with these parties. Since the composition and size of the corpus strongly influences the subsequent sentiment analysis, we suggest a party-independent corpus selection schema. In particular, we have decided to include the name of the political party and their youth organization, chairpersons, deputy chairpersons, secretary generals and honorary chairpersons of the parties, the names of the chancellors and ministers of the old and new governments, state premiers (Ministerpräsidenten) as well as the members of the federal parliament (Bundestag) during both legislative periods. We exclude members of state and communal parliaments or governments and ministerial staff (in particular secretaries of state) due to their low coverage in the chosen target media. This comprises our entity corpus.

Search Engine-Based URL Crawling

The first step of our data pipeline is to crawl URLs of articles from domains in our newspaper database. The goal is to find such articles that are *relevant* to the

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

public opinion-forming process and that have been read by a large number of people. Typical approaches employ graph-based spider algorithms that recursively search for links on a certain domain (various open-source tools exist such as (scrapy.org, 2018). However, this type of approach typically yields highly undirected and irrelevant results, in particular when one is interested in links from a single domain, as opposed to articles discussing a common topic. Therefore, we employ a search engine-based approach using the *Google Cloud Custom Search JSON API* (cloud.google.com, 2018). With a search engine, relevance is determined by the search engine's internal page rank algorithm, intuitively matching the attention an article might gain from a human reader.

Each search query consists of a search term, a target domain, the target language, and a time period of interest. The search term is an entity from a subset of the entity corpus (we limit the search terms to eliminate a large number of queries with no results). The target domain specifies the newspaper website from which to crawl articles. The target language is German in our use case. Furthermore, in order to have an article distribution that is homogeneous in publication date shifted over the full observation period. If queries were made for each search term and target domain during the whole observation period, the article distribution would be biased toward more recent articles, as a more recent publication date tends to increase the relevance of an article. We limit query results to such publications that lie in consecutive time periods, shifted over the full observation period. To determine the publication date at query time, we use *Page Dates*, which is metadata provided by the search engine that estimates the date of an HTML page based on the features from its content, such as dates in the title and URL. The number of URLs extracted in a specific time period varies greatly, due to the API being upper-bounded to ten URL queries. Therefore, we dynamically adjust this time period based on a protocol applied to the number of URLs extracted during the last query. If this number of URLs is high, the time period is shrunk so to limit the risk of exceeding the API upper bound with possibly relevant URLs to extract; if it is low, the time period is increased to avoid wasting quota.

URL Augmentation and Access

The URLs extracted by the search engine point to a particular page (e.g., but not always the first page) of an article. However, we are interested in the full text of a crawled article, in particular to provide the full semantic context for all entities given. Therefore, we build a domain-specific augmentation mechanism that appends a page suffix or infix in the URL crawled by the search engine and accesses it using the HTTP/HTTPS protocol. If the domain responds with an HTTP 404 error, the page does not exist. Otherwise, the accessed URL is added to the URL database. At the same time, we store the complete HTML pages of the accessed URLs in an intermediary storage.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Text and Metadata Extraction

Having crawled the URLs' HTML pages, we now aim at extracting selected contents from it. This is required, since the largest amount of information on a typical HTML page on the web contains cluttered strings, for example, for the layout or advertisements, which are excluded from the analysis. We extract the article's title, the paragraph content and metadata information (such as the publication date) using the open-source tool *beautiful soup* (crummy.com, 2018). This required us to manually find the domain-specific HTML tags for each content type. One problem we were confronted with was that on some domains, advertisements were embedded with the same HTML tags as paragraph content, and could therefore not be distinguished without further action. We solved this problem by finding characteristic, domain-specific identifiers that were—if contained in the extracted paragraph strings—used to exclude that part of the content.

Entity Extraction

In this fourth step of our data pipeline, we want to extract entities from the crawled article contents that correspond to the entities in our original entity corpus. Often, articles use variations of these corpus entities that refer to them. For example, the extracted entities "Sozialdemokratische Partei," "Sozialdemokratische Partei Deutschlands," and "Sozis" all map to the corpus entity "SPD." We therefore manually include common synonyms, colloquial terms, and abbreviations that can be directly mapped to entities in the corpus as illustrated in the example in Figure 2 . Aiming at a quantifiable approach, one may fall back to machine learning approaches such as *word2vec*, which classifies semantically similar words based on high-dimensional vector representations (Mikolov, Chen, Corrado, & Dean, 2013), or *wordNet*, which compares semantic similarities hierarchically (Miller, 1995). One may also parse the newspaper articles in search of party-associated keywords by exploiting a multilingual entity database, for example, as established by Al-Rfou, Kulkarni, Perozzi, & Skiena, 2015).

Entity Sentiment Analysis

Given the extracted entities, we now aim at calculating sentiments. The term *sentiment* is defined as "settled opinion reflective of one's feelings" (Pang et al., 2008). Sentiment analysis, which has been studied extensively in both academia and industry, aims at analyzing whether an entire paragraph or article expresses a positive or negative sentiment. While the literature differentiates multiple types of sentiments and related emotions, in this study, we focus on the simple classification of positive and negative sentiment on a continuous, upper- and lower-bounded scale. Sentiment is specified as a decimal number ranging between −1 (negative) and 1 (positive). Naive sentiment analysis calculates the sentiment over an entire sentence and lumps together all words. As displayed in Figure 3, this may lead to a positively depicted entity being associated with a negative sentiment. To tackle this

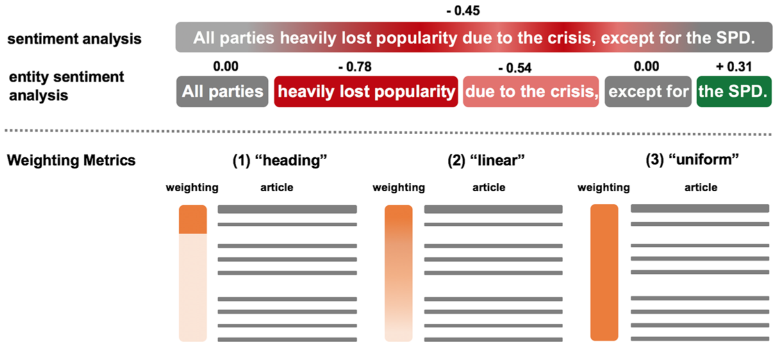

Figure 3. Schematic Illustration of the Entity Sentiment Analysis.

problem, it is important to distinguish sentiment analysis from entity sentiment analysis. *Entity sentiment analysis* is a more fine-grained technique that analyzes every word (entity) separately with respect to its semantic meaning in the sentence. For instance, the sentence "all parties heavily lost popularity due to the crisis, except for the SPD" contains words like "lost" and "crisis" and may therefore be associated with a negative overall sentiment. The noun "SPD", however, is clearly separated from these negative entities and assumes a positive sentiment when considered on an entity level. The IBM natural language processing API and similar platforms offer fee-based entity sentiment analysis services. Additionally, Chen and Skiena provide a multilingual, open-source implementation based on knowledge graph propagation algorithms and embeddings trained on the Wikipedia corpus (Chen & Skiena, 2014). By utilizing their pretrained implementation, we not only manage to keep our entire working pipeline open-source, but also achieve a greater flexibility in the implementation.

More than ever before, in the era of fast-paced news sharing on social media, people tend to read the headlines and first paragraphs rather than whole articles. Taking this into account, we implement three different metrics as shown in the bottom half of Figure 3. Each metric weights the parts of an article differently, as indicated by color strength: (i) "headline"—the title and subtitle are allocated a strong weighting making up to 80 percent irrespective of the article length; (ii) "linear"—from title to end of the article, the weighting linearly decreases; and (iii) "uniform"—arguing that we want to analyze the relationship between newspapers and political parties, we may decide to disregard any paragraph weighting and rather focus on the entire article. The selection of the appropriate weighting metric is solely based on the question of whether news reports are in mutual interplay with their readers or stand by themselves.

For comparison with the above outlined approach, we use the IBM Watson Natural Language Understanding API for entity sentiment extraction, which performs steps 4 and 5 of the data pipeline in a blackbox-like fashion (ibm.com, 2018). The API applies a uniform weighting metric to the entities.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Hierarchical Entity Matching

Since entity sentiment analysis is working with unstructured text data, there exists a large variance of observed raw entity names referring to the same politician (e.g., “Frau Merkel” and “Chancellor Merkel,” both referring to the entity “Angela Merkel”). Figure 4 shows various examples of raw entities yielded from the entity sentiment analysis associated with the two politicians Angela Merkel and Martin Schulz, chancellor candidates of the two largest political parties during the 2017 election. Thus, observations need to be matched to the targeted entities in our corpus.

We use a *hierarchical* matching approach consisting of the following steps to map the raw entities to exactly one entity in our corpus. First, observed entities are checked if they directly match the search term of the article. Otherwise, the raw entity name is compared with all search terms, including only the important politicians. Lastly, if the entity could still not be directly mapped, the name is compared with all remaining entities in the corpus. For all three steps, a *full match* applies, if the raw entity contains the first and last name of the politician in the corpus. As this is not always applicable, *surname matches* are also included. If several surname matches were found, we use the *Levenshtein distance* (Miller, Vandome, & McBrewster, 2009) to determine the match corresponding to the highest similarity between the raw entity and the corpus entity. The measurement is often used in fuzzy string searching and for personal name matching (Christen, 2006). To ensure high data quality of surname matches, they are verified manually. If entity names reveal noisy matches, such as for common surnames like “Müller” or “Maier”, such observations are discarded. In recent years, performance in named entity recognition has improved significantly as a result of adopting machine learning based, neuronal architectures (Lample, Ballesteros, Subramanian, Kawakami, & Dyer, 2016). In principle, the matching procedure could be replaced by training such an architecture that directly recognizes and identifies an entity from a predefined set of politicians. However, this is overloading this subtask in terms of complexity and would in particular require extensive labeled training data. In a final cleaning step, we exclude from

Figure 4. Word Clouds of Raw Entities Matched to the Politicians Angela Merkel and Martin Schulz, the Chancellor Candidates During the 2017 Federal Election.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2. The Newspaper Domains in the Final Data Set and Their Categorization, Together With the Numbers of Articles, Mentioned Politicians, and Entities.

Newspaper Domain	Category	Articles	Observations
bild.de	Newspaper	10,182	37,931
der-postillon.com	Satire newspaper	247	701
faz.net	Newspaper	13,585	50,683
fr.de	Political newspaper	9,645	36,348
focus.de	Political magazine	21,590	94,886
handelsblatt.de	Newspaper	5,569	17,417
huffingtonpost.de	Newspaper	7,520	33,871
jungefreiheit.de	Political newspaper	1,548	6,504
jungewelt.de	Political newspaper	2,802	6,764
n-tv.de	News medium	8,217	36,065
spiegel.de	Newspaper	11,084	56,815
stern.de	Political magazine	3,572	12,916
sueddeutsche.de	Newspaper	15,116	51,945
taz.de	Newspaper	10,712	39,443
tagesschau.de	Newspaper	3,024	14,245
tagesspiegel.de	Newspaper	8,360	34,942
welt.de	Newspaper	20,460	92,384
zeit.de	Newspaper	12,059	51,925
afdkompakt.de	Party newspaper (AfD)	2,758	15,861
bayernkurier.de	Party newspaper (CSU)	1609	6540
gruene.de	Party newspaper (Die Grünen)	217	750
national-zeitung.de	Party newspaper (DVU)	45	208
neues-deutschland.de	Party newspaper (Die Linke)	10,478	33,918
unsere-zeit.de	Party newspaper (DKP)	705	2,086
vorwaerts.de	Party newspaper (SPD)	1,934	6,970
Total	25 newspapers	183,038	742,118

our analysis infrequently mentioned politicians with fewer than 50 observations in the observation period. After data preprocessing and cleaning, 742,118 observations of 1,258,371 previously extracted observations remain. In table 2, the newspaper domains, their categorization by type and the number of crawled articles and observations in the cleaned data set is summarized.

Analysis—Political Media Sentiment in Germany 2017/2018

Having introduced the general framework, we turn toward a more detailed analysis of the data set. We applied our approach exclusively to German newspapers and articles in the German language for pragmatic reasons (the authors are domain experts in the German language and political landscape). However, note that our approach is generic and can be applied to the media landscape of any country, assuming the required tools are available for that language. We chose the observation period between January 2017 and April 2018 to surround the federal elections in Germany that took place in September 2017. Note that in this observation period, an exceptionally high number of five state elections and one federal convention (election of the federal president) took place. The observation period also comprises two legislative periods of the federal parliament. We chose

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

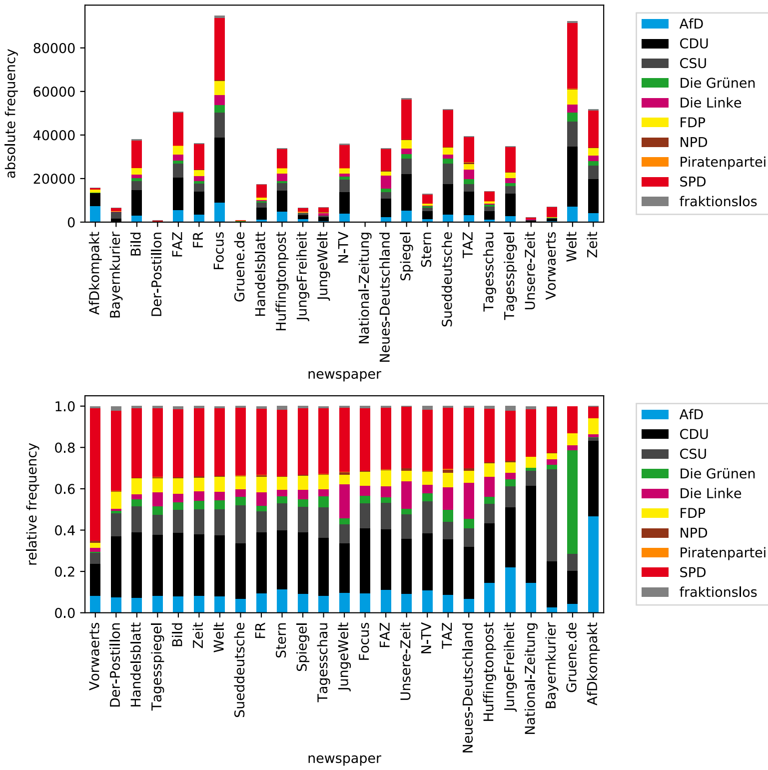

Figure 5. Top: Absolute Entity Mentions per Newspaper and Party. Bottom: Relative Entity Mentions per Newspaper and Party.

this observation period, since it corresponds to the most recent major political event in Germany, and is of high interest due to a six-month long negotiation and foundation period of the government until March 2018. All analytical results are based on sentiments extracted using the IBM NLP API in steps 4 and 5 of the data pipeline discussed in the previous section. We have published the final data set together with all data analysis provided in a user-friendly Python Jupyter Notebook on the website www.politicalcompass.de.

Descriptive Analysis

To get a first overview of the data set, we analyze the absolute and relative number of entities per newspaper and party affiliation in Figure 5. Two observations can be made: First, the total number of entities highly varies depending on the newspaper, as every newspaper publishes differing numbers of articles. In particular, party newspapers yield less coverage than commercial newspapers. Second, party newspapers such as AfDkompakt.de (AfD), Vorwaerts.de (SPD), Bayernkurier.de (CSU), and Gruene.de (Die Grünen) have significantly higher coverage of representatives from their own party in comparison with others. This result is expected: The deviation from the normal

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

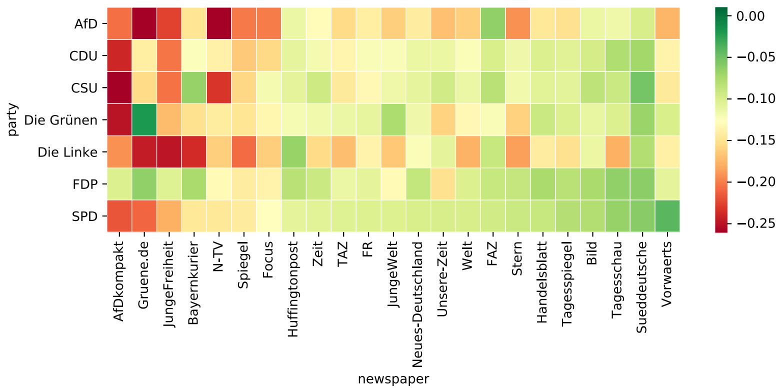

Figure 6. Sentiments Averaged Over all Entities of a Particular Party, Split by Newspaper.

relative report frequency indicates a political bias of such newspapers. This component of partisan bias is the coverage bias.

Party-Level Analysis. We begin by analyzing the sentiments on a party level: Figure 6 analyzes the average sentiment over all entities of a political party separately for each newspaper. Several interesting conclusions can be drawn. First, the governing parties (CDU, CSU, and SPD) are rather positively evaluated over almost all newspapers. Second, parties of the right and left, like the AfD and Die Linke, are rather negatively evaluated by all newspapers. In particular, the AfD, which in 2017 entered the federal parliament for the first time, is covered very negatively in accordance with the public opinion that media would negatively cover this party. Nearly, fifteen out of seventeen political magazines and newspapers in Figure 6 write most negatively about either the AfD or Die Linke with the only exceptions being FAZ and the Huffington Post. This implies that the general media is very critical toward these two parties. On the contrary, it gives the political parties the opportunity to place themselves in a victim role and to utilize this position to attract "anti-establishment" voters. "Lüegenpresse" ("lying press") is a commonly used exclamation at AfD demonstrations. Entity sentiment analysis is not able to verify so-called "fake news," but certainly shows evidence of a more negative attitude toward the AfD and its members by the established media, a phenomenon known as "AfD bashing." At the same time, the analysis shows that AfDkompakt, the party newspaper of the AfD, reports most negatively about other parties, expressing more discontent than any other media outlet. These findings indicate a trend of political division in the German discourse, which is increasingly dominated by emotions.

Third, sentiments of the party newspapers of Die Grünen, SPD, and CSU show a particularly positive average sentiment toward their own party, strengthening the argument that entity sentiment analysis is a valid method for exposing political bias. Surprisingly, the level of sentiment of AfDkompakt toward AfD representatives is relatively low compared with FDP. A negative bias toward its own party is rather unlikely. One explanation could be innerparty disunity. The fact that AfDkompakt

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

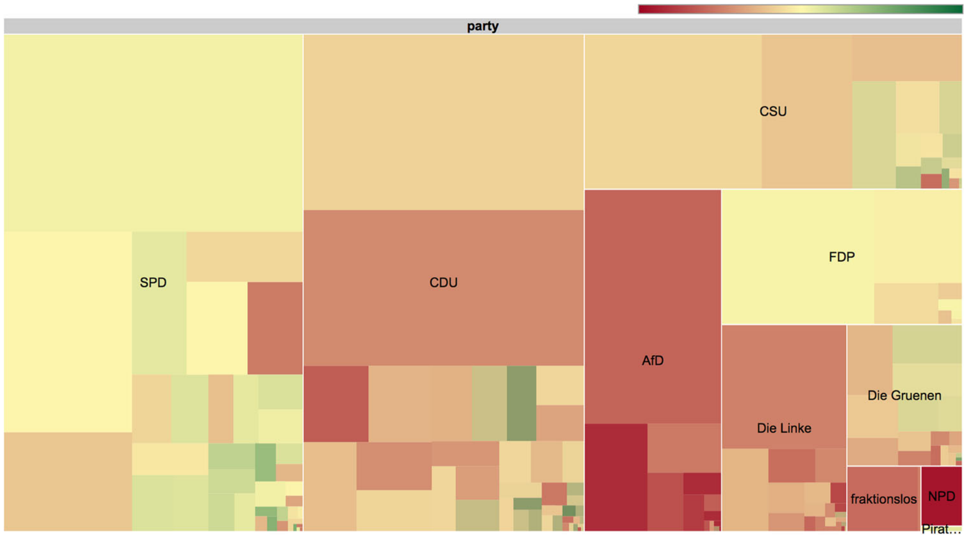

Figure 7. Heat Map Split by Political Party (Areas) and Entities (Subareas). The Size of the Areas Corresponds to the Number of Sentiments Found in Articles. The Color Reflects Comparatively Negative (Red) or Positive (Green) Sentiments.

generally yields more negative sentiments about all parties supports their characterization as a protest party (Arab, 2017). Given this general negativity of articles, misleading sentiments could be resulting for AfD politicians, if they are mentioned in an overall negative article and the sentiment extracting algorithm is directing this global judgement to all entities in the article—including any AfD politicians mentioned.

Entity-Level Analysis. After analyzing the newspapers' coverage at a party level, we now dive a level deeper and analyze the sentiments toward politicians. Figure 7 shows a hierarchical tree map representing the sentiments of a political party. The bigger the party's subarea in the map, the more articles mention the party and its affiliated politicians. Every political party can be subdivided into smaller tiles that reflect mentions of individual politicians of the respective party. We find varying sizes of the tiles according to the number of mentions. A red-colored tile indicates a negative sentiment, and a green color a comparatively positive sentiment. As red tiles are predominant, we may conclude that the media sentiment is generally more negative than positive, coinciding with the common allegation against newspapers. Moreover, we find that some parties are more heterogeneous than others, in the sense that the sentiments toward their affiliated politicians exhibit a wider range. While all politicians of the AfD party are reported homogeneously negatively, the politicians of the CDU party range from strongly negative (dark red) to strongly positive (dark green).

In order to understand the sentiment discrepancies further, Figure 8 shows the distribution of negative and positive sentiments, ranging from −1 to 1 for the CDU politicians Thomas de Maizière and Daniel Günther.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

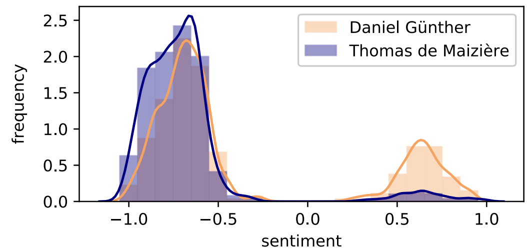

Figure 8. Histogram of Positive Versus Negative Mentions of Thomas de Maizière and Daniel Günther.

The imbalance of negative compared with positive sentiments underlines the generally critical reporting toward politicians. Moreover, Günther is mentioned more frequently in a positive context. He was elected as state premier (Ministerpräsident) of Schleswig-Holstein in June 2017.

Transforming the real-valued sentiments into the discrete, binary categories "positive" and "negative," a comparison of the relative frequencies in Figure 9 yields a

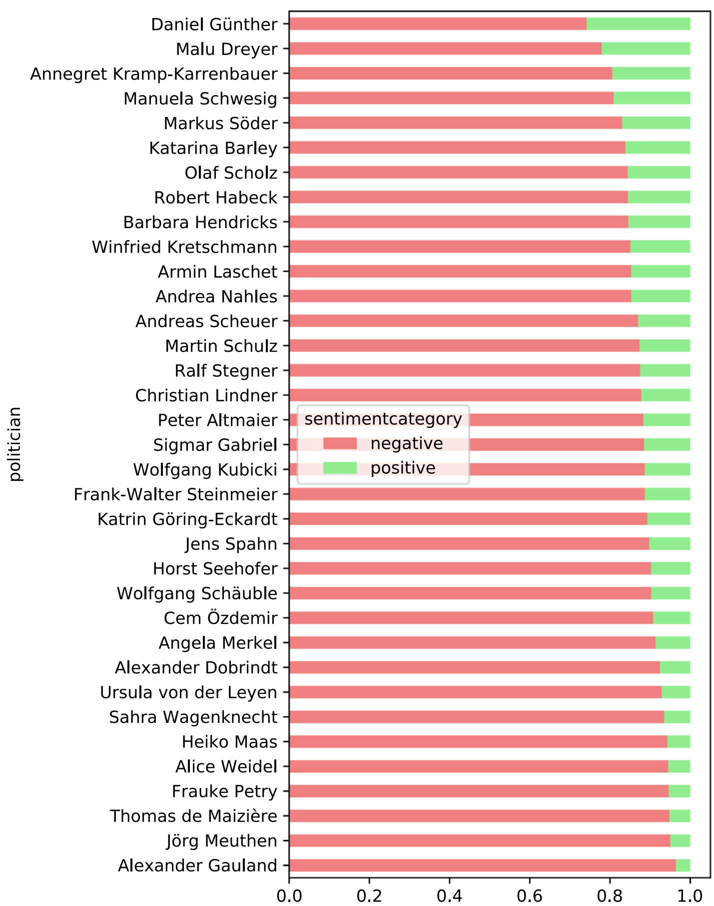

Figure 9. Comparison of Positive Versus Negative Mentions of Several Important Politicians.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

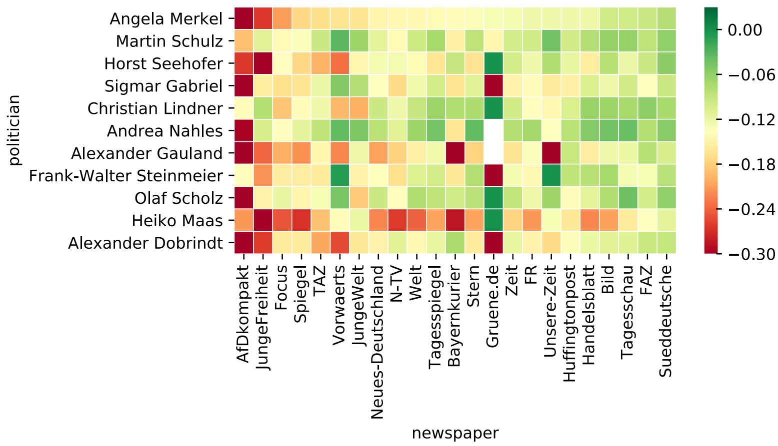

Figure 10. Sentiments Averaged Over Selected entities, Split by Newspaper.

ranking reflecting popularity and media acceptance of the politicians. Four of the five most positively ranked politicians (Daniel Günther, Annegret Kramp-Karrenbauer, Manuela Schwesig, and Markus Söder) were elected or promoted to state premier (Ministerpräsident) of a federal state of Germany during the analyzed time frame. This is another strong indication of the influence of newspaper bias on public opinion.

A heat map of the average sentiments of politicians in different newspapers in Figure 10 further examines the bias discovered at the politician level. Naturally, the SPD party newspaper Vorwärts positively mentions Martin Schulz, Sigmar Gabriel, Andrea Nahles, Frank-Walter Steinmeier, and Olaf Scholz, all politicians of the SPD party, while the same politicians are criticized in AfDkompakt. In comparison, in Junge Freiheit, a pattern of positive attitude toward FDP politician Christian Lindner and several SPD politicians is visible, suggesting a socioliberal tendency of the newspaper.

Next, we bring together three perspectives: the party-level, the politician-level, and the newspaper-level views. The Sankey diagram in Figure 11 visualizes the number of mentions grouped according to these three aggregates. We observe the number of sentiment occurrences in newspapers and political parties (left → middle) and the number of sentiments affiliated with parties split by their corresponding politicians (middle → right). Two observations are particularly striking: First, the number of sentiments split by the different parties in the middle section is roughly proportional to the share of politicians of that party in the federal parliament. A second interpretation results from considering the right half of the Sankey diagram: A few entities make up most of the sentiments of a political party. For example, the entities CDU and Angela Merkel make up roughly half of the sentiments associated with the CDU party. This result is very surprising, as roughly one hundred entities were used as search terms in the data crawling stage. Consequently, only a few politicians influence the political opinion, as they are overproportionally covered by news media.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

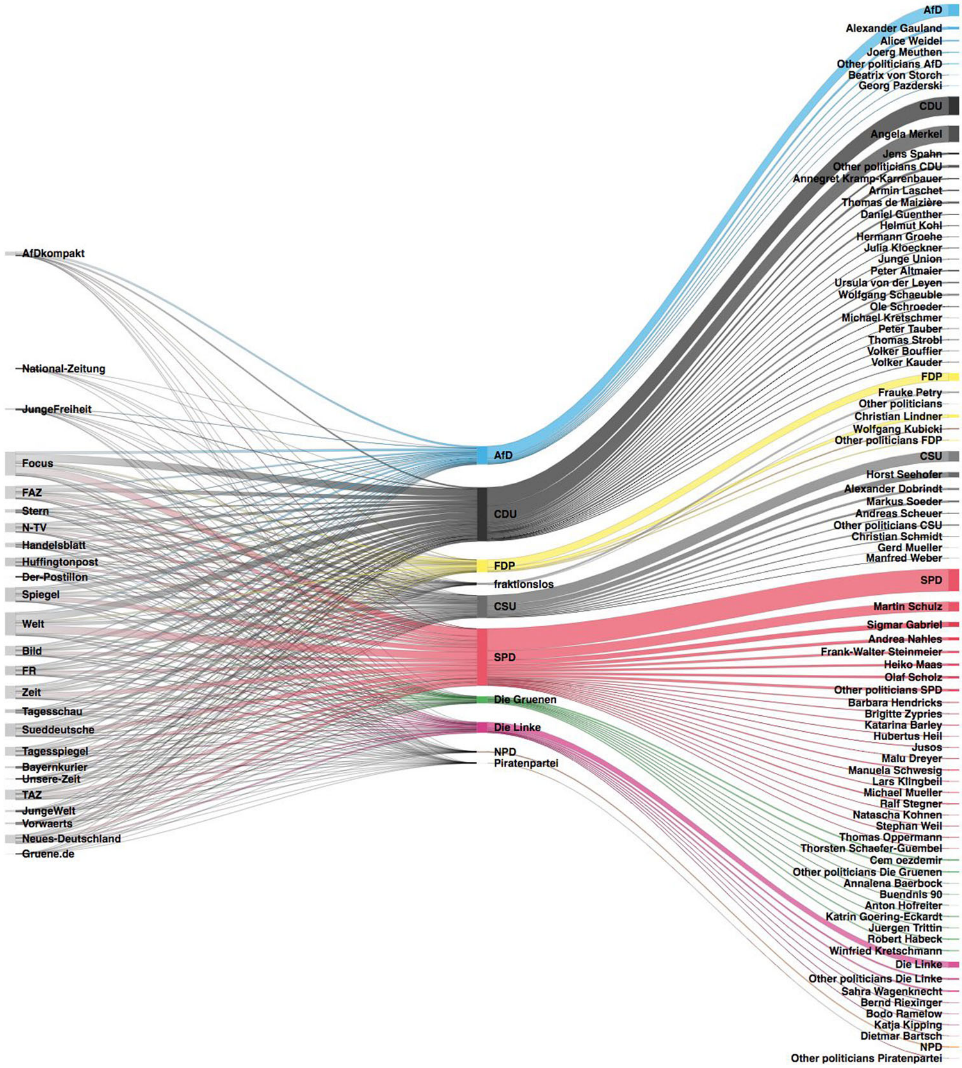

Figure 11. Sankey Diagram Visualizing the Number of Sentiments Occuring in Newspapers That are Affiliated With Political Parties (Left → Middle) and the Party's Split by Entity (Middle → Right).

Temporal Analysis. As a last step of the descriptive analysis, we turn toward the temporal dimension. Figure 12 visualizes the total number of sentiments per week, as extracted from the articles with a publication date in a certain week. We can observe that the number of sentiments is *not* distributed uniformly over the course of time. There exist several peaks next to important events such as elections or negotiations and rather small numbers of sentiments extracted during the summer slump (June to August) and the winter holiday (December and January). This effect was expected, since important political events produce more relevant political news, which is subsequently reflected in the articles crawled by the search engine.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

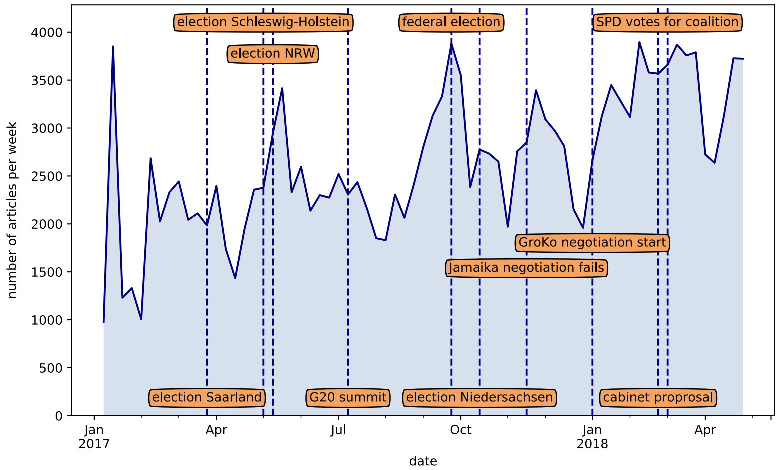

Figure 12. Absolute Number of Sentiments per Week. Markers Indicate Important Political Events That Might Have an Impact on the Sentiment Trajectory.

The sentiment of a politician or party over time is a noisy signal with high variance as seen before in the distribution in Figure 8 resulting from multiple negative and positive observations in any chosen time frame. This is an undesirable property of the time-series sentiments, as it prevents the analysis of smooth trends in the data. One way to treat this would be to use temporal aggregation (e.g., daily, weekly, or monthly). However, this would increase the influence of periods with fewer mentions, and would lead to heteroscedastic aggregations that oscillate in periods with fewer mentions. Instead, we treat the time series as a sequence and use convolution with window functions to smooth the signal. In particular, the Boxcar (moving average) is often used for trend identification in time series problems in general, and for sentiment time series in particular (Giachanou & Crestani, 2016). We apply an extension of the moving average principle called the *Bohman window* (Kim & Park, 2010). The only difference to the moving average is that observations more distant in terms of the time dimension from the current time step, for which a smoothed sentiment value shall be computed, are weighted less compared with close observations. Therefore, the smoothed sentiment on a given day is more influenced by sentiments from the same day compared with sentiments a week before. We employ smoothing with a weighted window to reflect event-related, sudden changes in sentiments. Given an input sentiment series x including all observed sentiments of an entity ordered by time, the general Bohman window with size M is defined as

$$W_M(x) = \begin{cases} \left(1 - \frac{|x|}{M}\right)\cos\left(\pi\frac{|x|}{M}\right) & \text{if} \quad |x| \leq M, \\ \quad + \frac{1}{\pi}\sin\left(\pi\frac{|x|}{M}\right) & \\ 0 & \text{if} \quad |x| > M \end{cases}$$

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Formally, a sentiment series S^P of a politician in the observed time $t \in 1,\dots,T$ with $t = 1$ referring to the first and T referring to the last observed time step can be formulated as follows:

$$S = \{S_1, S_2, \dots, S_T\}$$
$$S_t = \{s_{t,1}, s_{t,2}, \dots,\ s_{t,n_t}\}$$

The ith sentiment at time step t of a politician is described by $s_{t,i}$. Note that the number of observations n_t of a politician varies from day to day. To smooth the sentiment, we define the size M of the window depending on the average number of observations in a given time interval Δt. For example, if sentiments are smoothed on a two-week basis and a politician has two hundred observations on average in all two-week periods in the time series, we chose a window size of two hundred. To formally calculate the window size, one must introduce an additional index $j \in \{1,\dots,J\}$ with every j representing one time interval and $J = \frac{T}{\Delta t}$. The window size is then defined as

$$M = \max\left\{\frac{1}{J}\sum_{j=1}^{J} C(j, \Delta t),\ 3\right\}$$

The max-operator limits the minimum size of the window to three in the case of very few sentiments. $C(j, \Delta t)$ denotes a counter function of observations in a given time interval, defined as

$$C(j, \Delta t) = |\{s_{t,i} \mid \forall i \in \{1,\dots,n_t\}, \forall\ t \in [j \cdot \Delta t, (j+1) \cdot \Delta t]\}|$$

Finally, the denoised sentiment series of an entity is extracted by convolving the window through the whole series $S = \{s_{1,1},\dots,s_{1,n_1}, s_{2,1},\dots,s_{2,n_2}, s_{3,1},\dots,s_{T,n_T}\}$ with x indexing all observations from the original entity sentiment time series as

$$\tilde{S} = (S*W)(x) = \sum_{k=1}^{M} S(x-k)W(k)$$

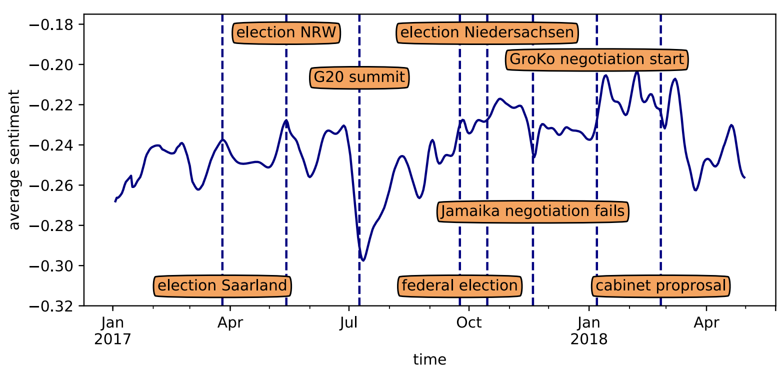

Figure 13. Average Denoised Sentiment Over Time.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

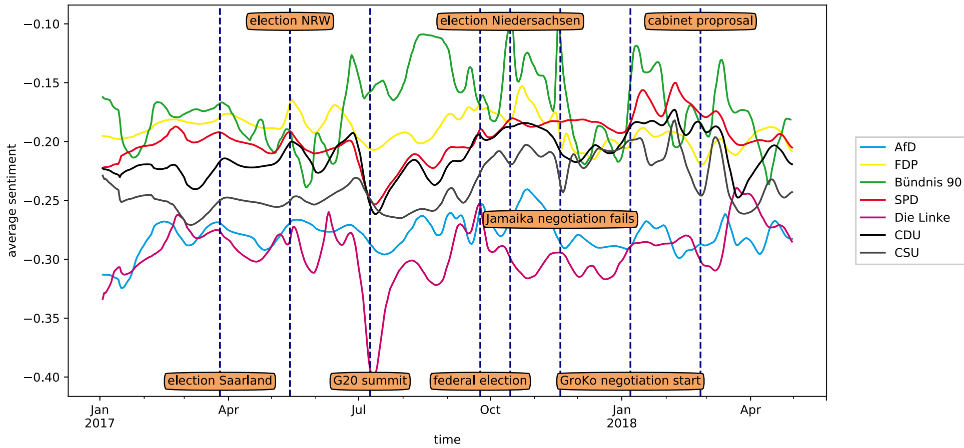

Figure 14. Average Sentiment by Parties Over Time.

Defining the window as dependent on the number of observations in a time window is intuitive, since it will yield time-consistent, smoothed sentiments for politicians, independent of their mention frequency in the media. This makes smoothed sentiments of different entities time-comparable. In our case, we set the time interval Δt to four weeks, providing a good trade-off between a denoised, smoothed time series while conserving changes in the sentiment.

Figure 13 illustrates the temporal change of average, denoised sentiments over time. We observe a correlation between major political events (annotated in the figure) and the average news sentiment. An eye-catching anomaly can be observed in July 2017, around the G20 summit that took place in Hamburg, Germany. The sentiment in all newspapers is more negative in this period. One explanation could be the civil violence and demonstrations against the summit that were strictly condemned by public opinion.

Figure 14 sheds another light on this finding as it separates the aggregated sentiment by political party. The media sentiment toward the left-wing party Die Linke is particularly negative in this time period. This may be traced back to the allegations that most of the violent protestants were left-wing extremists and accusations that the party did not clearly enough dissociate itself from the violent acts. To give another example of the influence of a political event on public sentiment, in the context of the so-called "Jamaika coalitition negotiations," the four involved parties FDP, Die Grünen, CDU, and CSU experience a drop in the sentiment. After a four-week long process, the liberal party FDP unexpectedly terminated the negotiations and was criticized for this act, reflected in a downward peak and decreasing trend thereafter in its sentiment time series.

Figure 15 reveals sentiment changes over time for the main representatives of the seven largest political parties. For example, in January 2017, Sahra Wagenknecht (Die Linke, a left-wing party) was confronted with major criticism from within her party after attacking the refugee policy of Angela Merkel by using an argument that had previously been voiced by right-wing parties.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

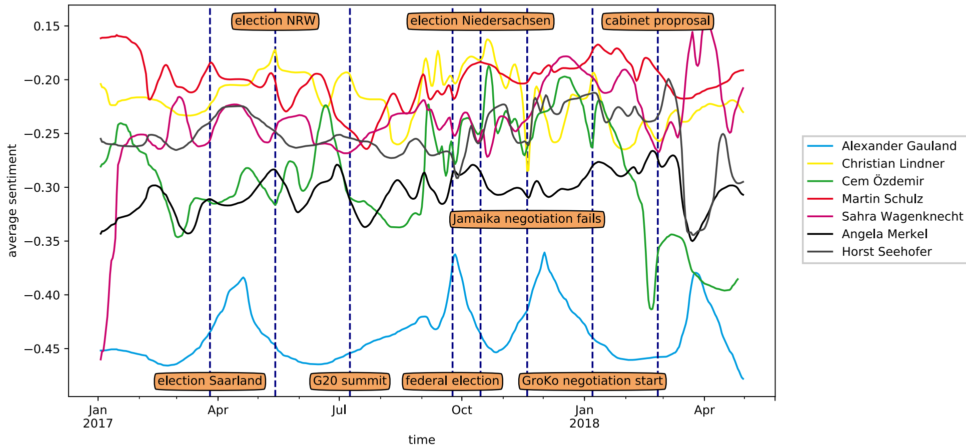

Figure 15. Average Sentiment by Politicians Over Time.

Another clear change in sentiment can be observed for Cem Özdemir in the beginning of 2018 before his resignation as federal chairman of Die Grünen. Lastly, the average sentiment of Horst Seehofer (CSU party chairman) decreased in March 2018 after stating his famous sentence "*Der Islam gehört nicht zu Deutschland*" (*Islam does not belong to Germany*.). This started a debate between Angela Merkel and Horst Seehofer, leading to a more negative coverage for both politicians. There are many more changes visible for further research on this data set.

As a last step toward the Sentiment Political Compass, we assess the time-consistency of political biases by grouping the average temporal sentiments by newspaper. In Figure 16, the sentiment of Martin Schulz, the SPD's candidate for chancellor, is compared between Focus (blue), the SPD party newspaper Vorwärts (red), as well as the average of all other newspapers (yellow). The latter can be interpreted as a "general opinion," as it reflects the sentiment across all other media outlets. First, Focus tends to write less positively about Martin Schulz and first reports with an above-average sentiment after his resignation. Second, Vorwärts

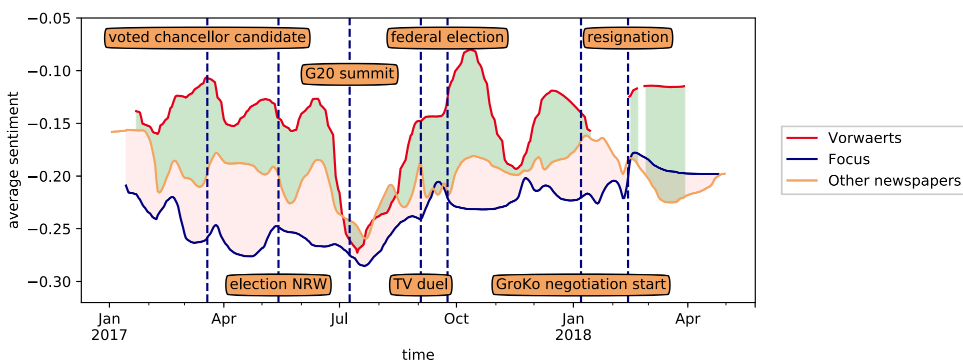

Figure 16. Average Sentiment of Martin Schulz by Groups Over Time.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

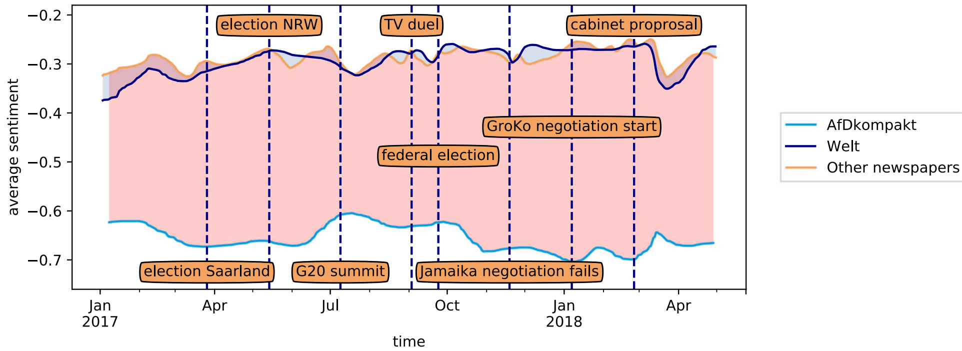

Figure 17. Average Sentiment of Angela Merkel by Groups Over Time.

reports more positively compared with the average opinion, with the only exception being after the aforementioned G20 summit. Additionally, we observe that there were no mentions of Schulz during the intraparty discussions before his resignation in February 2018.

In Figure 17, we similarly compare Angela Merkel's sentiment grouped by Welt, AfDkompakt and all other newspapers. Here, the sentiment of Welt toward Angela Merkel is not significantly different compared with the general opinion in other newspapers. However, the sentiment of Angela Merkel in AfDkompakt is strongly negative throughout the analyzed time frame, suggesting a long-term political bias and showing its measurability with entity sentiments.

Sentiment Political Compass

Recalling our results in the previous section, we already gained detailed insights into the newspapers' conviction toward political parties and their affiliated politicians. The previous analysis focused on a fine-grained view of the sentiments of each newspaper toward parties and politicians separately. With the *Sentiment Political Compass* (*SPC*), we aim at combining these proximities of newspapers toward all parties in a well-established framework.

As with its predecessors, the SPC is a two-dimensional figure. It aims at classifying political attitudes on two scales: left versus right, and libertarian versus autocratic. In our specific case, we make use of the political classification of political parties in the SPC by politicalcompass.org and extend it with the positions of *newspapers* toward these parties (politicalcompass.org, 2001). Our goal is to find a two-dimensional embedding of the newspapers' conviction relative to fixed party locations that as closely as possible represents the underlying biases and positions in the data. Contrary to prior approaches, the SPC is data-driven based on entity sentiments, making it reproducible with machine-based methods.

We model the political compass within a two-dimensional space $\mathbb{R}^2$. Each party is given at a fixed location (δ_1, δ_2). For each party, our data yields distributions of sentiments toward entities associated with a party, which vary by newspaper.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Formally speaking, let S be a continuous random variable in the interval $(-1, 1)$, N be a newspaper, and Y be a political party. Then, we are given distributions of the sentiments conditional on newspaper and party $P(S \mid N, Y)$. Each conditional distribution is well-defined and constituted by a statistically sufficient, yet potentially small number of observations in our data set. To increase the number of samples per conditional distribution, we use a technique inspired by statistical *bootstrapping* that samples from the distribution with replacement. In the Sentiment Political Compass, these conditional distributions represent the sentiment of the newspaper at the exact point of a particular party location, assuming that they are given as "ground truth."

All that said, the fixed party locations are assumed to be *uncertain*: The positions of political parties within the applied scales is to some degree subjective and imprecise, as its measurement relies on a questionnaire-based evaluation. Furthermore, the positions of political parties are time-dependent. Since we are only given the positions of the political parties once for the whole observation period, it is likely that they change over time and adapt due to new political circumstances and events. Therefore, instead of considering samples from the conditional distribution at the exact party location, we utilize *jittering*, which introduces an artificial variance circular around the party location and expresses this uncertainty.

With this technique highly resembling *Monte Carlo sampling*, we now draw a total of one million jittered sentiment samples from our conditional distributions at all party locations. By forming discrete bins for every region in the political compass, we can extract the mean bias of the newspaper at each location.

Specifically, for each bootstrapped sample, we draw jitter samples γ_1, γ_2 for the two coordinates from a standard normal distribution

$$\gamma_1, \gamma_2 \sim \mathcal{N}(0, 1)$$

and weight them by multiplication with a positive hyperparameter α. Note that $\gamma_i \alpha$, $i \in \{1, 2\}$ can be negative. α is user-defined depending on the level of uncertainty toward the party locations that shall be expressed. Also, α is equal for all newspapers and parties. We apply the jittering and shift every bootstrapped sample of the sentiment distribution away from the party location by adding the weighted jittering sample to both coordinates of the sentiment sample as

$$\delta_1^* = \delta_1 + \gamma_1 \alpha$$

$$\delta_2^* = \delta_2 + \gamma_2 \alpha$$

δ_1^* and δ_2^* refer to the new coordinates of the bootstrapped sentiment sample after jittering.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Figure 18. Sentiment Distribution of Several Newspapers Using Jittered, Bootstrapped Samples From Their Distributions Conditional on Party and Newspaper. The Figures Illustrate the Bias of Newspapers in the Political Dimensions Left vs. Right and Libertarian vs. Autocratic.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

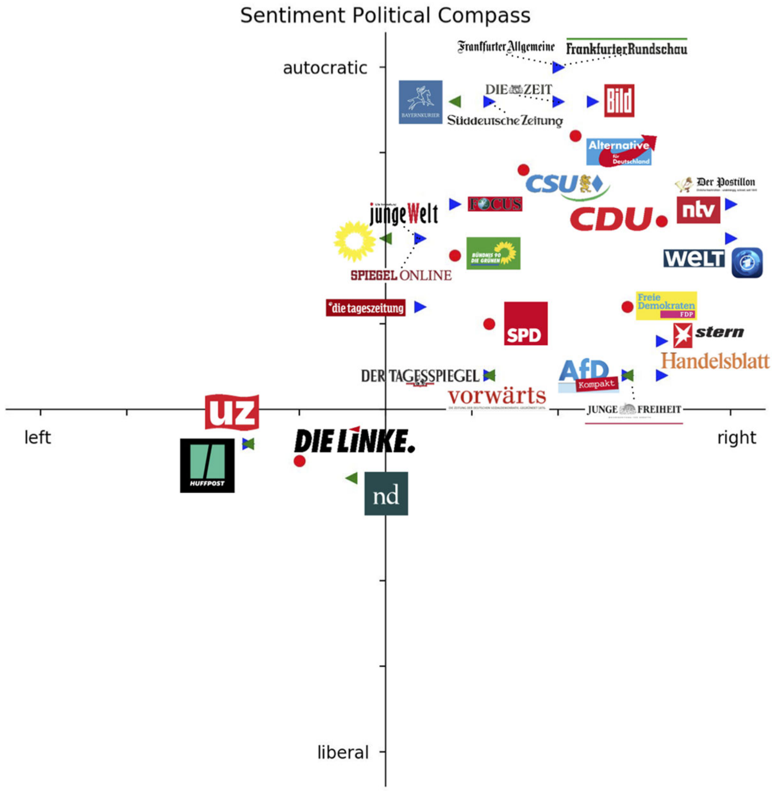

Figure 19. The Sentiment Political Compass. The Party Locations (Red Points) Are Fixed, the Party-Affiliated Newspaper Locations (Green Triangles) and all Other Newspaper Locations (Blue Triangles) Are Estimated With Their Corresponding Bias Maps in Figure 18.

As a last step, we aim at reducing the fine-grained distribution analysis by newspaper to a point estimate of the newspaper, similar to the given point locations of the political parties. To do so, we follow the intuition that the location of a newspaper should be defined as the center of a region with the most positive bias. Formally, this position is extracted with the argument of the maximum over all means of every 3×3 block of bins containing information on political bias illustrated in Figure 18. With these point estimates from the sentiment distributions in Figure 18 for each newspaper, we obtain the final product, the Sentiment Political Compass, in Figure 19.

The results for a few example newspapers are illustrated in Figure 18. Note that far autocratic-left and far liberal-right drawn sentiments are missing, as there are currently no German parties positioned in these regions and thus no information on political bias exists. The upper four maps show the political bias of party newspapers. As expected, each of them displays a stronger (positive) political bias when political attitudes are close to the associated party. For example, Bayernkurier has an autocratic bias while strongly rejecting the left-liberal quadrant. Vorwärts positions itself closer to left-liberal than autocratic-right and Unsere Zeit is pulled

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

toward left-liberal believers. Considering the newspapers Welt and Focus, there is no high gradient in terms of sentiment. Only the newspaper Welt has a small negative bias toward the left-liberal. In contrast, the newspaper Junge Freiheit shows a stronger positive political bias in the liberal-right quadrant claimed by the FDP. One very strong negative bias is observed for N-TV toward autocratic parties, like the AfD and CSU. Since CNN was the major shareholder of the media channel from its launch in 1992 until 2002 (SpiegelOnline, 1992), further research could compare political tendencies across international networks. Moreover, a qualitative study on articles confirms this bias, with frequent mentions of negative connotations and contexts of the word "autocratic" (N-TV, 2018).

Several additional insights can be deduced from the Sentiment Political Compass. First, the party newspaper Vorwärts is in proximity to SPD, Neues-Deutschland is in proximity to Die Linke, Bayernkurier is close to CSU, and Gruene.de also shows proximity to the location of Die Grünen. With these findings, the sanity check for the Compass that party newspapers tend to be close to their corresponding parties is passed. Only AfDkompakt misses its spot next to AfD, due to the misleading negative sentiments toward its own party discussed above. Second, for all other newspapers we can identify five clusters in close proximity to parties: One cluster consisting of Unsere Zeit and the Huffington Post is located in the left-liberal quadrant next to Die Linke. Unsere Zeit is the party newspaper of the German Communist Party, again validating the approach. A survey conducted in 2014 also demonstrated the left-leaning tendencies of readers of the US-based Huffington Post (Pew Research Center, 2014). One cluster of newspapers in close proximity to the FDP consists of Junge Freiheit, Stern, and Handelsblatt. One cluster close to Die Grünen consists of Focus, Spiegel, Junge Welt, and TAZ. A cluster of newspapers that seems to write less negatively about autocratic parties like the AfD and CSU compared with other newspapers consists of FAZ, FR, Süddeutsche, Bild, and Zeit. Less autocratic, but further in the neo-liberal right, the last cluster consists of Tagesschau, N-TV, and Welt and is in close proximity to the CDU.

Conclusion

In this article, we have presented the Sentiment Political Compass, a data-driven framework to analyze political bias in the online media landscape. We addressed the research problem by developing and outlining a technical framework to crawl articles, and perform entity extraction and entity sentiment analysis. The framework is generic to the degree that it can be reproduced in any other context that involves textual content in the political domain, including other countries, or social media and print media (or indeed, video content, where transcripts can be obtained). Furthermore, we showcased the analytical power of the Sentiment Political Compass by considering online news from twenty-five publishing houses in Germany during the election year of 2017/2018. In addition to the two-dimensional compass itself, we provide a detailed descriptive analysis of sentiment trends at the party and entity level, and the

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

time-variant effects in particular with respect to major political events. The insights are striking, as we, for example, confirm negative media bias toward the populist party AfD, which at the same time expresses strongly negative sentiment toward other political parties through AfDkompakt, reflecting a political division in society. On the contrary, we find that the amount of media coverage per party nearly corresponded to the distribution of votes. This contrasts with the common belief that a few "protest parties" like the AfD tend to dominate all media reporting.

In future work, we would like to consider entities other than politicians and political parties, such as labor unions, nongovernmental organizations, supranational or religious associations and companies. Furthermore, by collecting metadata on article types, one could group the database by editorials and news articles, so that a drilled-down analysis of the composition of partisan bias from coverage and sentiment bias would be feasible. Further extension by the authors will include real-time updates to monitor the latest news and trends. The media sources may also be extended to social networks and print media. In addition, we must analyze the degree to which the construct "newspaper" is the result of an institutional voice that is actively fostered within a publishing house. Through these efforts, the Sentiment Political Compass aims at exploiting the possibilities of data science to contribute to a more fact-based and informed political discourse.

Being politically neutral, a main motivation behind our work is to tackle misinformation and deficiencies in media reporting with transparency and quantifiability. Media bias may affect uninformed voters in democratic opinion-forming. Holding quantifiable evidence, however, problems may be soundly discussed on a scientific basis. In addition, empowering editorial staff with these insights could potentially lead to more balanced reporting. Utilizing the Sentiment Political Compass, measures can be taken to inform, monitor, or even regulate search engines or news tagging. The data and the source code of our framework are publicly available on www.politicalcompass.de, to support the hope that our approach will be applied to different countries and problems on a global scale.

Fabian Falck, Department of Computing, Imperial College London, London, UK [fabian.falck17@imperial.ac.uk].

Julian Marstaller, Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Karlsruhe, Germany.

Niklas Stoehr, Department of Computer Science, University of Oxford, Oxford, UK.

Sören Maucher, Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Karlsruhe, Germany.

Jeana Ren, Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Karlsruhe, Germany.

Andreas Thalhammer, Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Karlsruhe, Germany.

Achim Rettinger, Institute of Computational Linguistics and Digital Humanities, Trier University, Trier, Germany.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Rudi Studer, Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, Karlsruhe, Germany.

References

Al-Rfou, R., V. Kulkarni, B. Perozzi, and S. Skiena. 2015. "Polyglot-NER: Massive Multilingual Named Entity Recognition." In *Proceedings of the 2015 SIAM International Conference on Data Mining, Vancouver, British Columbia, Canada, April 30–May 2, 2015*.

Anstead, N., and B. O'Loughlin. 2014. "Social Media Analysis and Public Opinion: The 2010 UK General Election." *Journal of Computer-Mediated Communication* 20 (2): 204–20.

Arab, A. 2017. *Welche Parteien die meisten Stimmen an die AfD verloren (Which parties have lost the most votes to the AfD)*. https://www.welt.de/politik/deutschland/article168989573/Welche-Parteien-die-meisten-Stimmen-an-die-AfD-verloren.html

Beck, K. 2012. *Grundlagen und Infrastrukturen des deutschen Mediensystems (Foundations and infrastructures of the German media system)*, VS Verlag für Sozialwissenschaften, Wiesbaden, 33–72.

Bessi, A., M. Coletto, G.A. Davidescu, A. Scala, G. Caldarelli, and W. Quattrociocchi. 2015. "Science vs Conspiracy: Collective Narratives in the Age of Misinformation." *PLOS ONE* 10 (2): 1–17.

Boecking, B., M. Hall, and J. Schneider. 2015. "Event Prediction With Learning Algorithms—A Study of Events Surrounding the Egyptian Revolution of 2011 on the Basis of Micro Blog Data." *Policy & Internet* 7 (2): 159–84.

Butenschön, R., D. Ganser, J. Wernicke, G. Hautsch, D. Goeßmann, U. Krüger, D. Dahn, K. Gritsch, K.-J. Bruder, J. Becker, et al. 2017. *Lügen die Medien?: Propaganda, Rudeljournalismus und der Kampf um die öffentliche Meinung (Does the media lie?: propaganda, swarm journalism and the fight for public opinion)*. Westend Verlag.

Carbonell, J.G. 1979. "Subjective Understanding: Computer Models of Belief Systems." Technical Report, Yale University, Department of Computer Science.

Caton, S., M Hall, and C. Weinhardt. 2015. "How Do Politicians Use Facebook? An Applied Social Observatory." *Big Data Society* 2 (2): 1–18.

Ceron, A., L. Curini, S.M. Iacus, and G. Porro. 2014. "Every Tweet Counts? How Sentiment Analysis of Social Media Can Improve Our Knowledge of Citizens Political Preferences With an Application to Italy and France." *New Media & Society* 16 (2): 340–58.

Chen, Y., and S. Skiena. 2014. "Building Sentiment Lexicons for all Major Languages." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. New York, NY: ACM, 109–17.

Christen, P. 2006. "A Comparison of Personal Name Matching: Techniques and Practical Issues." In *Sixth IEEE International Conference on Data Mining—Workshops (ICDMW'06)*, 290–94.

Chung, J., and E. Mustafaraj. 2011. "Can Collective Sentiment Expressed on Twitter Predict Political Elections?" In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, 1.

cloud.google.com. 2018. Google Cloud—Custom Search JSON API. https://developers.google.com/custom-search/json-api/v1/overview

crummy.com. 2018. *Beautiful Soup Documentation*. https://www.crummy.com/software/BeautifulSoup/bs4/doc/

De Fortuny, E.J., T. De Smedt, D. Martens, and W. Daelemans. 2012. "Media Coverage in Times of Political Crisis: A Text Mining Approach." *Expert Systems With Applications* 39 (14): 11616–22.

Diener, E. 1980. "Deindividuation: The Absence of Self-Awareness and Self-Regulation in Group Members." In *The Psychology of Group Influence*, ed. Paulus, P.B. Hillsdale: L. Erlbaum, 209–42.

Eysenck, H.J. 1957. *Sense and Nonsense in Psychology*. Oxford, England: Penguin Books, Inc.

Federal Centre for Political Education of Germany. 2002. *Wahl-o-mat—Q&A Tool to Evaluate the Proximity of Personal Political Conviction to Political Parties*. Federal Centre for Political Education of Germany.

Ferguson, L.W. 1941. "The Stability of the Primary Social Attitudes: I. Religionism and Humanitarianism." *The Journal of Psychology* 12 (2): 283–88.

Fritz, M. 1987. *World's Smallest Political Quiz*.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Gaebler, B. 2017. *AfD und Medien—Analyse und Handreichungen (AfD and the Media—Analysis and Recommendations).*

Gao, W., and F. Sebastiani. 2015. "Tweet Sentiment: From Classification to Quantification." In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.*

Gautam, G., and D. Yadav. 2014. "Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis." In *2014 Seventh International Conference on Contemporary Computing.*

Gayo-Avello, D. 2013. "A Meta-Analysis of State-of-the-Art Electoral Prediction From Twitter Data." *Social Science Computer Review* 31 (6): 649–79.

German Audit Bureau of Circulation. 2018a. German Audit Bureau of Circulation—Goals & Tasks. http://www.ivw.de/ivw/ziele-und-aufgaben-ivw

German Audit Bureau of Circulation. 2018b. *IVW—Definitions and Technical Explanations.* http://daten.ivw.eu/download/pdf/Online_RichtlinienV2_1_Anlage1.pdf

German Audit Bureau of Circulation. 2018c. *IWF—Online Offerings—Pageimpressions.* http://ausweisung.ivw-online.de/index.php

Giachanou, A., and F. Crestani. 2016. "Tracking Sentiment by Time Series Analysis." In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, 1037–40.

Greenberg, J., and E. Jonas. 2003. "Psychological Motives and Political Orientation—The Left, the Right, and the Rigid: Comment on jost et al." *Psychological Bulletin* 129 (3): 376–82.

Haselmayer, M., M. Wagner, and T.M. Meyer. 2017. "Partisan Bias in Message Selection: Media Gatekeeping of Party Press Releases." *Political Communication* 34 (3): 367–84.

Holtzman, N.S., J.P. Schott, M.N. Jones, D.A. Balota, and T. Yarkoni. 2011. "Exploring Media Bias With Semantic Analysis Tools: Validation of the Contrast Analysis of Semantic Similarity (cass)." *Behavior Research Methods* 43 (1): 193–200.

Hurlimann, M., B. Davis, K. Cortis, A. Freitas, S. Handschuh, and S. Fernandez. 2016. "A Twitter Sentiment Gold Standard for the Brexit Referendum." In *Proceedings of the 12th International Conference on Semantic Systems.*

ibm.com. 2018. *IBM Watson—Natural Language Understanding API.* https://www.ibm.com/watson/services/natural-language-understanding

Janis, I.L. 2008. "Groupthink." *IEEE Engineering Management Review* 36 (1): 36.

Jungherr, A., P. Jürgens, and H. Schön. 2011. "Why the Pirate Party Won the German Election of 2009 or the Trouble With Predictions." *Social Science Computer Review* 30 (2): 229–34.

Jungherr, A., H. Schön, and P. Jürgens. 2013. "The Mediation of Politics Through Twitter: An Analysis of Messages Posted During the Campaign for the German Federal Election 2013." *Journal of Computer-Mediated Communication* 21 (1): 50–68.

Kaczmirek, L., P. Mayr, R. Vatrapu, A. Bleier, M. Blumenberg, T. Gummer, A. Hussain, K. Kinder-kurlanda, K. Manshaei, M. Thamm, K. Weller, and C. Wolf. 2014. *Social Media Monitoring of the Campaigns for the 2013 German Bundestag Elections on Facebook and Twitter.* Mannheim: Gesis, Leibniz-Institut für Sozialwissenschaften.

Kim, H.-Y., and M. Park. 2010. "Classification of Time-Series Data Based on Several Lag Windows." *Communications of the Korean Statistical Society* 17 (3): 377–90.

Lample, G., M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer. 2016. "Neural Architectures for Named Entity Recognition." *arXiv preprint arXiv:1603.01360.*

Lester, J. 1994. "The Evolution of the Political Compass (and Why Libertarianism is Not Right-Wing)." *Journal of Social and Evolutionary Systems* 17 (3): 231–41.

LumenLearning. 2019. *Social Influence—Group Behavior.* https://courses.lumenlearning.com/boundless-psychology/chapter/social-influence/

Malouf, R., and T. Mullen. 2008. "Taking Sides: User Classification for Informal Online Political Discourse." *Internet Research* 18 (2): 177–90.

Marozzo, F., and A. Bessi. 2017. "Analyzing Polarization of Social Media Users and News Sites During Political Campaigns." *Social Network Analysis and Mining* 8 (1): 1–13.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Maute, K. 2011. *Das Mediensystem der USA und Deutschlands im Vergleich (The media system of USA and Germany in comparison)*. Macromedia Fachhochschule der Medien Stuttgart.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." *CoRR*, abs/1301.3781.

Miller, F.P., A.F. Vandome, and J. McBrewster. 2009. *Levenshtein Distance: Information Theory, Computer Science, String (Computer Science), String Metric, Levenshtein Distance, Spell Checker, Hamming Distance*. Alpha Press.

Miller, G.A. 1995. "WordNet: A Lexical Database for English." *Communications of the ACM* 38: 11.

Mutz, D.C. 1989. "The Influence of Perceptions of Media Influence: Third Person Effects and the Public Expression of Opinions." *International Journal of Public Opinion Research* 1 (1): 3–23.

N-TV (2011-2018). *Examples of Anti-Autocratic Articles*. https://www.n-tv.de/politik/dossier/Agypten-macht-die-Autokraten-nervoes-article25-95956.html, https://www.n-tv.de/politik/politik_person_der_woche/Dieser-Autokrat-ist-der-Gipfel-article204-76862.html, https://www.n-tv.de/politik/Europa-wird-autokratischer-article20348833.html, https://www.n-tv.de/mediathek/videos/politik/Autokraten-regieren-weltweit-mit-eiserner-Hand-article20552518.html

Neri, F., C. Aliprandi, F. Capeci, M. Cuadros, and T. By. 2012. "Sentiment Analysis on Social Media." In *Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*.

Nolan, D. 1969. "Classifying and Analyzing Politico-Economic Systems." *The Individualist (Society for Individual Liberty)*.

Pang, B., and L. Lee, et al. 2008. "Opinion Mining and Sentiment Analysis." *Foundations and Trends in Information Retrieval* 2 (1–2): 1–135.

Pang, B., L. Lee, and S. Vaithyanathan. 2002. "Thumbs Up?: Sentiment Classification Using Machine Learning Techniques." In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10*. Association for Computational Linguistics.

Pew Research Center. 2014. *Ideological Placement of Each Source's Audience*. http://www.pewresearch.org/pj_14-10-21_mediapolarization-08-2/

Pla, F., and L.F. Hurtado 2014. "Political tendency identification in twitter using sentiment analysis techniques." In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 1, 183–92.

politicalcompass.org. 2001. *The Political Compass*.

Ravi, K., and V. Ravi. 2015. "A Survey on Opinion Mining and Sentiment Analysis: Tasks, Approaches and Applications." *Knowledge-Based Systems* 89: 14–46.

scrapy.org. 2018. *Scrapy—Open Source and Collaborative Framework for Crawling Websites*. https://scrapy.org/

Silverthorne, C.P. 1971. "Information Input and the Group Shift Phenomenon in Risk Taking." *Journal of Personality and Social Psychology* 20 (3): 456.

SpiegelOnline. 1992. *Reich und berühmt (Rich and Famous)*. http://www.spiegel.de/spiegel/print/d-13682966.html

statista.com. 2018. *Net Reach of Top 15 Online Newspapers in Unique Users, as of May 2018*. https://de.statista.com/statistik/studie/id/6551/dokument/zeitungen-statista-dossier/

Stier, S., A. Bleier, M. Bonart, F. Mörsheim, M. Bohlouli, M. Nizhegorodov, and S. Staab. 2018. *Systematically Monitoring Social Media: The Case of the German Federal Election 2017*, 4. Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences.

Tumasjan, A., T. Sprenger, P. Sandner, and I. Welpe. 2010. "Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment." *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 178–85.

Twitter, Inc. 2019. *Twitter Developer API*. https://developer.twitter.com/en/docs

Wagner, J.-P. 2018. "Local Newspapers Coverage of Scottish Independence During the Campaign of the 2014 Scottish Independence Referendum." *Dealing With the Local*.

Wilks, Y., and J. Bien. 1983. "Beliefs, Points of View, and Multiple Environments." *Cognitive Science* 7 (2): 95–119.

19442866, 2020, 3, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/poi3.222 by Linkoping Universitet, Wiley Online Library on [03/10/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License