Topic and tone of political news articles in German online media.

Franziska Löw

April 10, 2018

Abstract

The aim of this paper is to investigate whether the political reporting of different content providers distinguishes itself and whether this reporting has an influence on the opinion-forming process of the voters.

Keywords Structural Topic Model, Sentiment Analysis, Unsupervised Machine Learning, Text Mining

Contents

1 Introduction

Pluralism of media is an essential principle of democratic societies. For the process of forming opinions, information conveyed by the media, in particular the mass media of journalism, plays a decisive role. They create the preconditions for making the social and cultural diversity of modern societies visible and manageable. For this reason, securing the diversity of opinion-forming processes and preventing concentration processes in the media sector is an essential prerequisite for a functioning democracy. An important question is therefore whether the political reporting of different content providers distinguishes itself and whether this reporting has an influence on the opinion-forming process of the voters. The aim of this paper is to investigate these questions with the use of up-to-date text mining methods. The analysis is based on online news articles from leading German news websites before, during and after the German parliamentary elections in September 2017. The internet as a source of information is gaining in importance, especially in a political context. According to a study on the media coverage of the German-speaking population, nearly 40% of respondents used the internet at least once a week in 2017 to inform themselves about the current news compared to 34% in 2016 (Medienanalyse), 2017).

The new developments in digital media provide new opportunities and challenges for media outlets. The costs of providing and disseminating information have been reduced and the speed of information exchange has increased enormously. At first glance, these developments seem to have a positive effect on the provision of information. However, this also means that the incentive for media to produce primary information declines since the Internet has reduced the marginal utility of content. A study of French news websites from 2013 shows that only 38 percent of the online articles contained original content, with the remainder being copies of those originals (Cagé, Hervé, and Viaud, 2017). A multitude of factors influence the diversity of online media, including the market structure and the increasing digitization of the industry. Media outlets (TV, Radio, Newspaper) usually operate in two-sided markets, where they serve two customer groups (content consumer and advertiser) that are connected via indirect network effects. The two-sided market structure of the private news market results in news platforms striving to choose their content in such a way that its reach is as large as possible in order to maximize profits from advertising revenues. Steiner (1952) concluded, that profit-maximizing media owners may choose to offer the same content, i.e. content aligned with the tastes of the majority. Gabszewicz, Laussel, and Sonnac (2001) study the problem of diversity of the political content of newspapers. They find that the maximum differentiation only prevails if the readers sufficiently value the political differentiation between the newspapers the advertising market is small enough. On the other hand, advertising may also have a positive impact on the media, as it enables publishers to report independently of political parties. Ellman and Germano (2009) analyze a theoretical framework of a two-sided market for newspaper, where readers value accuracy and advertisers value advert-receptive readers. They found that advertising increases the intensity of competition for readers and therefore raises accuracy of media coverage, whereas in the monopolistic case, newspapers under-report news that sufficiently reduces advertiser profits.

Recognizing the importance of media pluralism, the European Commission has endorsed several initiatives to discuss and promote media diversity in the European Union[1]. One way to ensure the independence and diversity of the media landscape is through public funded media. In Germany public broadcasting originated in the post-war period and is financed by compulsory fees. To take into account the distinct nature of digital media, the Interstate Broadcasting Agreement (Rundfunkstaatsvertrag) also regulates the scope for action of online services offered by public service broadcasting since 2007. Accordingly, public media are not allowed to distribute purchased content and must - depending on the category of content - set a time limit on its accessibility. In addition, there is a strict advertising ban and prohibition of regional reporting.

With the purpose of examining the media pluralism in the market for German-language online news, this paper analyzes German online news articles about domestic politics. German federal elections took place on 24th of September 2017 and the formation of the government has taken up a period of about five month. The articles considered here dated from 01.06.2017 to 01.03.2018 and thus inform their readers both about the election promises of the parties (before the election) and about the coalition talks (after the election). They therefore make an important contribution to the public's opinion-forming process. The empirical strategy follows a novel approach combining "the two Ts": Topic and Tone (Hansen and McMahon, 2016). This means that the topics discussed in the articles are identified first (topic), followed by an analysis of how they are discussed with the various news websites (tone). The final step is to check whether the tonality of reporting is reflected in the survey results on voting preferences calculating the cross correlation coefficients. More precisely, the research strategy is as follows:

1. **Discovering Topics (Section 5.1)** To discover the latent topics in the corpus, the structural topic modeling (STM) developed by M. E. Roberts, B. M. Stewart, and E. M. Airoldi (2016) is applied. The STM is an unsupervised machine learning approach that models topics as multinomial distributions of words and documents (as a synonym for news articles) as multinomial distributions of topics, allowing to incorporate external variables that effect both, topical content and topical prevalence. I estimate a model, where the news website is included as a control for both the topical content and the topical prevalence. The result of the generative process

[1] https://ec.europa.eu/digital-single-market/en/policies/media-freedom-and-pluralism

of STM are two posterior distributions: One for the topic prevalence in an article and one for the content of a topic. The latter is used to label the topics according to an event or issue discussed in the media. The topic prevalence of an article (or the posterior distribution) is used to assign a topic to each news article.

2. **Measuring Tone (Section 5.2)** After assigning a topic to each article, a dictionary-based sentiment analysis is conducted to estimate how topics are discussed differently by different news websites. The idea of a sentiment analysis is to determine the attitude of a writer toward the overall tonality of a document. To conduct such an analysis, a lists of words (dictionary) associated with a given emotion, such as negativity is pre-defined by the analyst. The document is then deconstructed into individual words and the frequencies of words contained in a given dictionary are calculated.

3. **Comparing with polls** In order to check whether the transmitted content from the online media is reflected in the voting preferences, the relationship between monthly average of the emotional value of individual topics from 2 and the survey value of a specific party is estimated using the cross correlation function (CCF).

Approach 1 has been used in M. E. Roberts, B. M. Stewart, and E. M. Airoldi (2016), among others, to examine differences in the content of eastern and western news providers regarding "the rise of China". However, I extend the analysis by comparing the sentiment score for a given topic an news provider, to identify which topics are discussed similar or differently (2). Hansen and McMahon (2016) applied a similar approach to a dataset of 142 FOMC (Federal Open Market Committee) decision statements to measure the effect of those statements on macroeconomic variables. An additional extension is the comparison with current election poll values (3).

The remaining course of the paper is as follows: The following Section 2 gives an overview about political trends in the past six month (June 2017 to March 2018). The data used to conduct the model is described in Section 3. Section 4 explains the generative process of the structural topic model as well as the selected parameters to run the model. The empirical analysis containing the above mentioned steps is conducted in Section 5.

2 Background on the federal election in Germany (2017)

The articles analyzed in this paper cover a period from June 1, 2017 to March 1, 2018 and thus cover both the most important election campaign topics for the Bundestag elections on September 24, 2017 and the process of forming a government that lasted until February 2018. After four years in a grand coalition with the Social Democrats (SPD), German Chancellor Angela Merkel, member of the Conservative Party (CDU/CSU, also known as Union), ran for re-election. The SPD nominated Martin Schulz as candidate for chancellor.

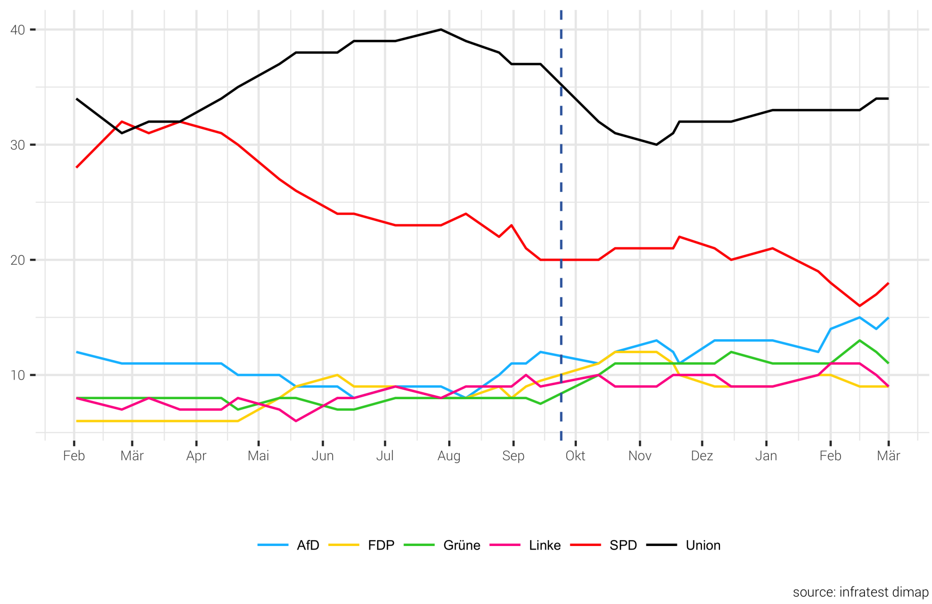
On the right side of the political spectrum, AfD (alternative for Germany) managed to be elected to the German Bundestag for the first time in 2017. The political debate about the high refugee numbers of the past years brought the AfD a political upswing by taking up the dissatisfaction of parts of the population and using it to raise its own profile. Leading party members of the AfD as well as party supporters repeatedly accused the mass media in the course of the reporting on the federal elections to report unilaterally and to present the AfD intentionally badly.

After the election, the formation of a government was difficult due to the large number of parties elected to the Bundestag and the considerable loss of votes by the major parties CDU/CSU and SPD. Since all parties rejected a coalition with the AfD, numerically only two coalitions with an absolute parliamentary majority were possible: a grand coalition ("GroKo" - from the German word Große Koalition) of CDU/CSU and SPD, and a Jamaica coalition (coalition of CDU/CSU, FDP (economic liberal party) and B90/Die Grünen (Bündnis 90/Die Grünen, green party)). The grand coalition was initially rejected by the SPD. The four-week exploratory talks on the possible formation of a Jamaica coalition officially failed on 19 November 2017 after the FDP announced its withdrawal from the negotiations. FDP party leader Christian Lindner said that there had been no trust between the parties during the negotiations. The main points of contention were climate and refugee policy. CDU and CSU regretted this result, while B90/Die Grünen sharply criticized the withdrawal of the Liberals. The then Green leader Cem Özdemir accused the FDP of lacking the will to reach an agreement.

After the failure of the Jamaica coalition, a possible re-election or a minority government as alternatives were discussed in the media before the SPD decided to hold coalition talks with the CDU/CSU. This led to great resistance from the party base, which called for a party-internal vote on a grand coalition. After the party members voted in favor of the grand coalition, a government was formed 171 days after the federal elections.

Figure 1 shows that support for the two major popular parties has been declining in recent months since August 2017, with the CDU/CSU again showing positive survey results since November. However, the value of the SPD has been falling since March 2017. At the same time, the AfD in particular has been recording increasingly positive survey results since June 2017. Section 5.3 examines whether there is a correlation between the survey results and the way the parties are reported in the media.

Figure 1: Election Polls



3 Dataset and data preparation

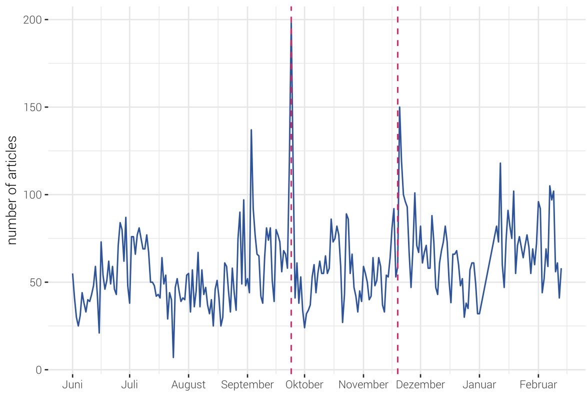
I conduct the estimation on a sample of 14,937 online news articles from seven german news provider about domestic politics[2]. The articles are dated from 01.06.2017 to

[2]Bild.de, DIE WELT, FOCUS ONLINE, SPIEGEL ONLINE, stern.de, ZEIT ONLINE, Tagesschau.de

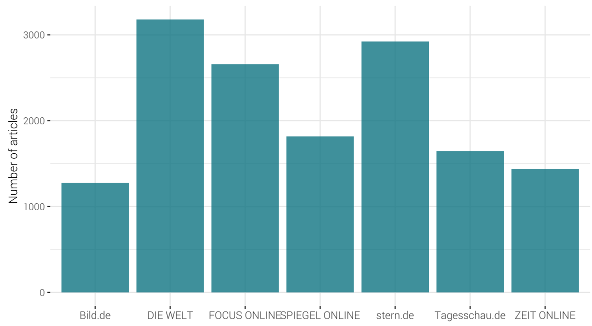
01.03.2018. I first extract all online articles using the the Webhose.io API.[3] Then all articles from the section "domestic policy" are filtered by using the URL of an article. Overall, the selected news providers are among the top ten German online news providers - in terms of unique user[4] - in the period under review, with only Tagesschau.de belonging to the public media. The reason for this is that the content structure of Tagesschau.de is most similar to that of the private providers. ZDF.de offers predominantly video content and DLF (Deutschlandfunk) website mainly offers audio content in the form of interviews, which makes it hard to include it in the model.

Figure 2a shows the distribution of the number of articles by date. There is a high peak around the federal elections on September, 24th and another one shortly after the failure of the Jamaica coalition talks on November, 19th (indicated by the red dotted lines). Figure 2b shows, that DIE WELT published most articles on domestic policy, followed by stern.de and FOCUS ONLINE.

Figure 2: Article distribution...



(a) ...by date



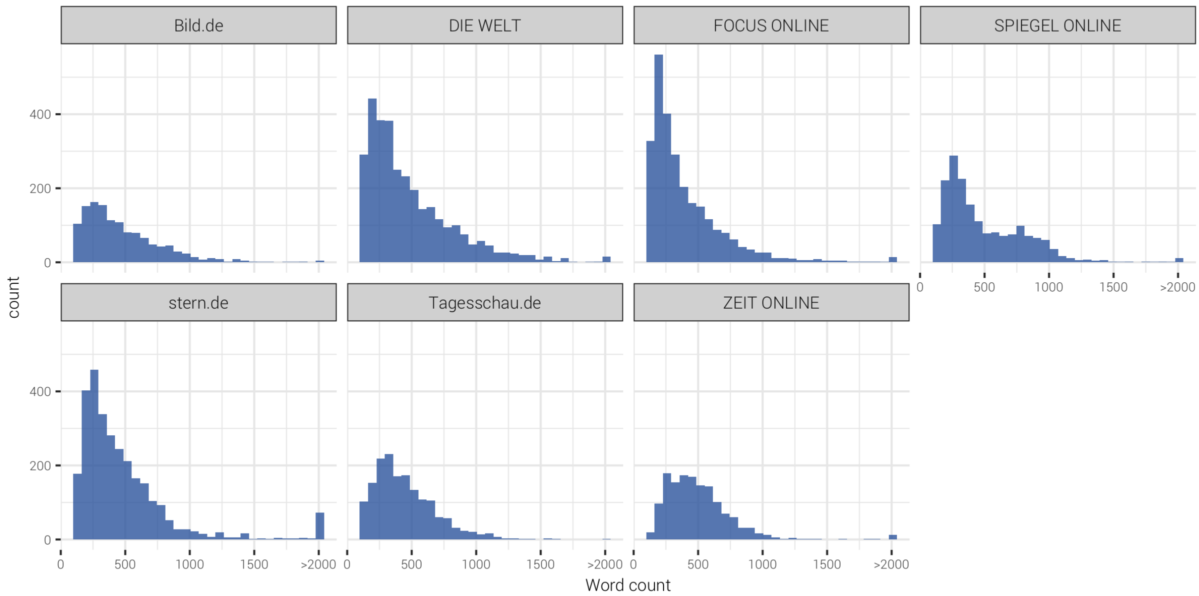
(b) ... by news source

Looking at the histograms of the word counts (Figure 3), it becomes evident that most of the articles have a word length between 200 and 1000 words (articles with less than 200 words were filtered out in advance, as these are mostly reader comments). The distribution at Bild.de, DIE WELT, FOCUS ONLINE and stern.de is left-skewed, whereby stern.de has comparatively many articles with a word length of over 2000.

[3]For more information see https://docs.webhose.io/v1.0/docs/getting-started. The scraping code was written in Python and can be made available on request.

[4]The term unique user refers to a number of different visitors to a website within a certain period of time. Multiple visits from the same user are only considered once.

Figure 3: Word Count



The longest article at Tagesschau.de has 2006 words, nevertheless the median value is comparatively large (see Table). ZEIT ONLINE has the highest median value of 459 words.

Table 1: Summary Statistic of Word Count

		n	mean	sd	median	min	max	se
1	Bild.de	1277.00	475.20	319.22	394.00	121.00	3710.00	8.93
2	DIE WELT	3179.00	507.88	614.28	377.00	121.00	14507.00	10.89
3	FOCUS ONLINE	2660.00	402.68	330.86	299.00	121.00	5647.00	6.42
4	SPIEGEL ONLINE	1817.00	498.96	333.23	387.00	121.00	3304.00	7.82
5	stern.de	2922.00	518.09	622.99	376.50	121.00	9287.00	11.53
6	Tagesschau.de	1644.00	450.34	242.93	397.50	121.00	2006.00	5.99
7	ZEIT ONLINE	1437.00	510.98	377.85	459.00	121.00	8015.00	9.97

To summarize the content of the texts, wordclouds help to get a first impression, as the frequency of words in a corpus are represented by their size. Intuitively the term frequency (tf) of a word is a measure of how important that word may be. The word cloud in Figure 4 is derived from all articles within the dataset. As can be seen, problems arise with words that are highly frequent. For example "die" ("the"), "und" ("and"), and "in" ("in") are extremely common but unrelated to the quantity of interest. These terms, often called stop words, are important to the grammatical structure of a text, but typically don't add any additional meaning and can therefore be neglected. A common strategy to reduce the number of language elements is to pre-process the text by imposing some preliminary restrictions (stop-word removal and stemming) based on the nature of the data (twitter text, newspaper articles, speeches, etc.) (Gentzkow, Kelly, and Taddy, 2017). In fact, to use text as data and reduce the dimensionality to avoid unnecessary computational complexity and overfitting, pre-processsing the text is a central task in text mining (Bholat et al., 2015).

Figure 4: Word Cloud (whole corpus)

Stemming is a process by which different morphological variants of a word are traced back to their common root. For example, "voting" and "vote" would be treated as two instances of the same token after the stemming process. There are many different techniques for the stemming process. I apply the widely used Porter-Stemmer algorithm, which is based on a set of shortening rules that are applied to a word until it has a minimum number of syllables.[5] To remove distorting words, the pre-defined stop word list from the Snowball project[6] together with a customized list of stop-words is used. Additionally punctuation character (e.g. ., „ !, ?, etc.) and all numbers are removed from our corpus. After completing this steps we were left with 68.576 unique terms in our vocabulary. The following wordclouds are derived from the corpus for each news provider. It becomes evident that these are texts discussing domestic policy issues. The SPD in particular seems to be highly frequent. However, at first glance, there are no obvious differences between the corpus of the different news provider.

To use the data for statistical analysis, the next step is to divide the whole corpus into individual documents and to represent these documents as a finite list of unique terms. In this setting, each news article represents a document d, whereby each of these documents can be assigned to a news website. Next, for each document $d \in \{1, ..., D\}$ the number of occurrences of term v in document d is computed, to obtain the count $x_{d,v}$, where each unique term in the corpus is indexed by some $v \in \{1, ..., V\}$ and where V is the number of unique terms. The D x V matrix $\boldsymbol{X}$ of all such counts is called the document-term matrix. This representation is often referred to as the bag of words model, since the order in which words are used within a document is disregarded. The sum of all documents

[5]https://tartarus.org/martin/PorterStemmer/

[6]http://snowball.tartarus.org/algorithms/german/stop.txt

forms what is called the corpus.

Figure 5: Word Clouds

(a) Bild.de

(b) DIE WELT

(c) FOCUS ONLINE

(d) SPIEGEL ONLINE

(e) Stern.de

(f) Tagesschau.de

(g) ZEIT ONLINE

4 The structural topic model

To find out the latent structure of each document, a structural topic model (STM) is estimated. In general, topic models formalize the idea that documents are formed by hidden variables (topics) that generate correlations among observed terms. They belong to the group of unsupervised generative models, meaning that the true attributes (topics) cannot be observed. One crucial assumption to be made for such models is the number of topics that occur over the entire corpus (K).

Each individual topic potentially contains all of the unique terms within the vocabulary V with different probability. Therefore, each topic k can be represented as a probability vector ϕ_k over all unique terms V. Simultaneously, each individual document d in the corpus can be represented as a probability distribution θ_d over the K topics. To generate each individual word $w_{d,n}$ in a document d for the n^{th} word-position, a topic allocation $z_{d,n}$ is drawn from the topic distribution for that document θ_d. Then, a word is drawn from the term distribution for the selected topic $\phi_{z_{d,n}}$.[7]

The STM developed by M. E. Roberts, B. M. Stewart, and E. M. Airoldi (2016) is a recent extension of the standard topic modeling technique, labeled as latent Dirichlet allocation (LDA), which refers to the Bayesian model in Blei, Ng, and Jordan (2003) that treats each word in a topic and each topic in a document as generated from a Dirichlet - distributed prior.[8] Since its introduction into text analysis, LDA has become hugely popular and especially useful in political science.[9] Wiedmann (2016) uses topic model methods on large amounts of news articles from two german newspapers published between 1959 and 2011, to reveal how democratic demarcation was performed in Germany over the past six decades. Paul (2017) compares editorial differences between media sources, using cross-collection latent Dirichlet allocation (ccLDA), an LDA-based approach that incorporates differences in document metadata. They use a dataset of 623 news articles from August 2008 from two American media outlets - msnbc.com and foxnews.com - to compare how they discuss topics. Reviewing the top words of the word-topic distribution, they find some content differences between the two media sources under review.

The STM allows to incorporate document specific covariates (e.g. the author or date of a document). The model has been applied to multiple academic fields: M. E. Roberts, B. M. Stewart, Tingley, et al. (2014) uses STM to analyze open-ended responses from surveys and experiments, Farrell (2016) applies the model to scientific texts on climate change, revealing links between corporate funding and the framing of scientific studies. Mishler et al. (2015) show that "STM can be used to detect significant events such as the downing of Malaysia Air Flight 17" when applied to twitter data. Another study shows how STM can be used to explore the main international development topics of countries' annual statements in the UN General Debate and examine the country-specific drivers of international development rhetoric (Baturo, Dasandi, and Mikhaylov, 2017). Mueller and Rauh (2016) use newspaper text to predict armed conflicts in different regions. They use the estimated topic shares in linear fixed effects regression to forecast conflict out-of-sample. M. Roberts, B. Stewart, and Tingley (2016a) use STM to examine the role of partisanship in topical coverage using a corpus of 13,246 posts that were written for 6 political blogs during the course of the 2008 U.S. presidential election. With the aim of revealing the effect of partisan membership on topic prevalence, each blog is assigned to be either liberal or conservative. To explore the differences between the two, they look at the expected proportion of topic and examine the posts most associated with a respective topic. This approach is similar to M. E. Roberts, B. M. Stewart, and E. M. Airoldi (2016).

4.1 Generative Process of STM

As mentioned above, the STM allows to incorporate observed document metadata which is able to affect both topical prevalence and topical content. These assumptions are reflected in the prior distributions. The following describes the generative process for filling the n^{th} word-position in document d in the case of the STM (M. Roberts, B. Stewart, Tingley,

[7]A more detailed description of the generative process of the STM can be found in Section 4.1

[8]See also Griffiths and Steyvers (2002), Griffiths and Steyvers (2004) and Hofmann (1999). Pritchard, Stephens, and Donnelly (2000) introduced the same model in genetics for factorizing gene expression as a function of latent populations.

[9]see Blei (2012), Grimmer and B. Stewart (2013) and Wiedmann (2016) for an overview in social science and Gentzkow, Kelly, and Taddy (2017) give an overview of text mining applications in economics.

and E. Airoldi, 2013): As in the case of conventional models, first a specific topic z_{dn} is assigned, according to the topic distribution for that document θ through the process:

$$z_{dn}|\theta_d \sim \text{Multinomial}(\theta_d). \tag{1}$$

To incorporate the covariate values for that document, a topic-prevalence vector θ_d is drawn from a logistic-normal distribution:

$$\theta_d|y_{d\gamma}, \Sigma \sim \text{LogisticNormal}(\mu = y_{d\gamma}\Sigma), \tag{2}$$

where $y_d\gamma$ lists the values of all metadata covariates for document d, where γ relates these covariate values to the topic-prevalence. The structure of Σ implies the possibility of correlations across documents in the topic-prevalence vector.

Conditional in the topic chosen (z_{dn}), a specific word w_{dn}, is chosen from the overall corpus vocabulary V, using the following process:

$$w_{dn}|z_{dn}, \phi_{dkv} \sim \text{Multinomial}(\phi_{dk1}, ..., \phi_{dkV}), \tag{3}$$

where the word probability ϕ_{dkv} is parameterized in terms of log-transformed rate deviations from the rates of a corpus-wide background distribution m_v (M. Roberts, B. Stewart, Tingley, and E. Airoldi, 2013). The log-transformed rate deviations can then be specified by a collection of parameters $\{\boldsymbol{\kappa}\}$, where $\kappa^{(t)}$ is a K-by-V matrix containing the log-transformed rate deviations for each topic k and term v, over the baseline log-transformed rate for term v. This matrix is the same for all A levels of covariates. To put it differently, $\kappa^{(t)}$ indicates the importance of the term v given topic k regardless of the covariates. Similarly, $\kappa^{(c)}$ is a A-by-V matrix, indicating the importance of the term v given the covariate level c regardless of the topic. Finally, $\kappa^{(i)}$ is a A-by-K-by-V matrix, collecting the covariate-topic effects:

$$\phi_{dkv}|z_{dn} = \frac{\exp(m_v + \kappa^{(t)}_{kv}, \kappa^{(c)}_{y_d v} + \kappa^{(i)}_{y_d kv})}{\sum_v \exp(m_v + \kappa^{(t)}_{kv}, \kappa^{(c)}_{y_d v} + \kappa^{(i)}_{y_d kv})}. \tag{4}$$

The STM maximizes the posterior likelihood that the observed data were generated by the above data-generating process using an iterative approximation-based variational expectation-maximization algorithm[10] available in R's stm package (M. Roberts, B. Stewart, and Tingley, 2016b).

This process generates two posterior distribution parameter:

1. ϕ is a K-by-V matrix (where K = number of topics and V = vocabulary), where the entry ϕ_{kvc} can be interpreted as the probability of observing the v-th word in topic k for the covariate level c (the news website).

2. θ is a D-by-V matrix (where D = number of documents and V = vocabulary) of the document-topic distributions, where the entry θ_{dk} can be interpreted as the proportion of words in document d which arise from topic k, or rather as the probability that document d deals about topic k.

In Section 5.1 θ is used to estimate the conditional expectation of topic prevalence for given document characteristics. In order to calculate the sentiment value in Section 5.2 each document d is assigned to the topic with the highest probability ($\max(\theta_k)$ for each document d).

[10] A technical description of this maximization process can be found in M. E. Roberts, B. M. Stewart, and E. M. Airoldi (2016)

4.2 Model and parameter selection

Inference of mixed-membership models, such as the one applied in this paper, has been a thread of research in applied statistics in the past few years (Blei, Ng, and Jordan, 2003) (Erosheva, Fienberg, and Lafferty, 2004) (Braun and McAuliffe, 2010). Topic models are usually imprecise as the function to be optimized has multiple modes, such that the model results can be sensitive to the starting values (e.g. the number of topics). Since an ex ante valuation of a model is hardly possible, I compute a variety of different models and compare their posterior probability. This enables me to check how results vary for different model solution (M. Roberts, B. Stewart, and Tingley, 2016a). I then cross-checked some subset of assigned topic distributions to evaluate whether the estimates align with the concept of interest (Gentzkow, Kelly, and Taddy, 2017). These manual audits are applied together with numeric optimization based on the topic coherence measure suggested by Mimno et al. (2011).

This process revealed that a model with 50 topics best reflects the structure in the corpus. Furthermore, the news website of the article is used as covariates in the topic prevalence and the topical content. In other words, the corresponding news website of an article influences the probability distribution of topics and how the topics are discussed.

5 Empirical Evaluation

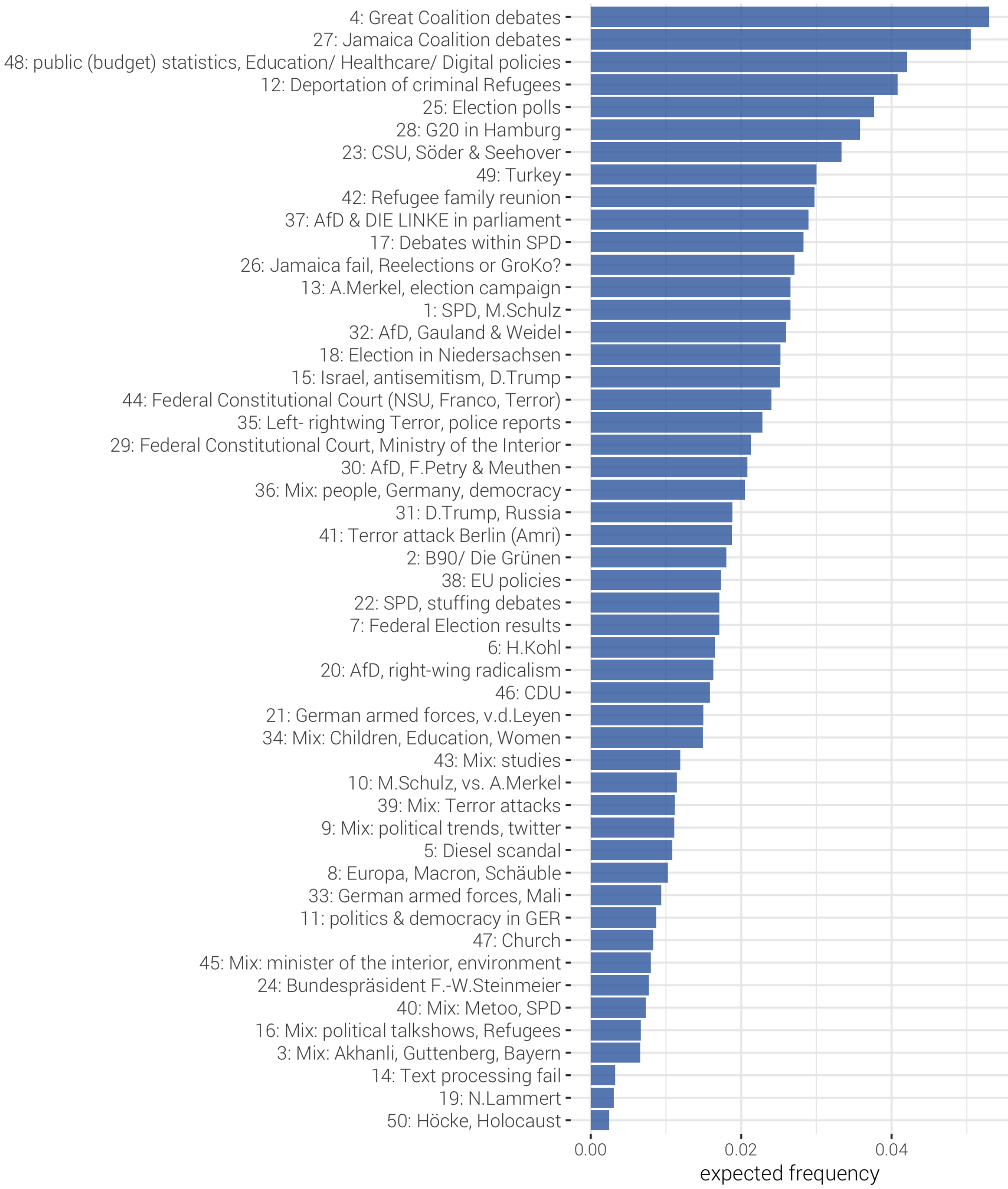
This section summarizes the results of the STM. Subsequently "the two T's" (Topic and Tone) of the corpus are analyzed according to the following approaches: (1) The document-topic probability θ_{dk} is used, to estimate the conditional expectation of topic prevalence for given document characteristics (See section 5.1). A set of topics is selected, that most distinctly discuss a particular party or a topic related to the federal elections. (2) Articles that are assigned to the selected topics with the highest probability are then used to conduct a dictionary-based analysis (see section 5.2). In order to check whether the sentiment value of certain topics are correlated with the results of voting preferences, the cross correlation function between these two concepts is calculated in 5.3.

5.1 Topic

In order to get an initial overview of the results, Figure 6 displays the topics ordered by their expected frequency across the corpus. To assign a label to each topic, I looked at the most frequent words in that topic and the most representative articles (M. E. Roberts, B. M. Stewart, and E. M. Airoldi, 2016).

It becomes apparent that topic 4 about the coalition talks between CDU/CSU and SPD - the "Grand coalition" or "GroKo" - is the topic with the highest expected frequency in the whole corpus, followed by the topic about the so-called Jamaica parties (CDU/CSU, FDP and B90/Die Grünen), which was the first alternative to be negotiated directly after the elections.

Figure 6: Expected topic proportion



The remaining analysis is limited to topics closely related to one or more parties. For this reason, the following topics were selected. The most frequent words of these topics at each news website can be seen in Appendix A.1:

1. Topic 1: About the SPD, mainly about the election campaign and Martin Schulz as candidate for the chancellor.

2. Topic 2: About B90/Die Grünen, mainly covering issues regarding the party's personell debates.

3. Topic 4: Covering the debates about the great coalition talks, mainly after the failure of the Jamaica coalition talks.

4. Topic 13: About Angela Merkel, mainly right before the election.

5. Topic 17: Covering votes within the SPD, mainly regarding the vote about a possible coalition with CDU/CSU/CSU ("GroKo").

6. Topic 20: About the AfD, mainly about their relation to right-wing extremist groups.

7. Topic 22: About SPD, mainly covering issues regarding the party's personell debates

8. Topic 23: About issues regarding the CSU, mainly about the competition between Horst Seehofer and Markus Söder and the negotiations with the CDU/CSU.

9. Topic 26: Discussing the failure of the Jamaica coalition talks and the two possible alternatives: Reelections or a great coalition.

10. Topic 27: Covering the Jamaica coalition talks, mainly focusing on the smaller players Bündnis B90/Die Grünen and FDP.

11. Topic 30: About the AfD, mainly about the resignation of Frauke Petry and Jörg Meuthen.

12. Topic 32: About the AfD, mainly about Alice Weidel and Alexander Gauland, voted as parliamentary party leaders after the resignation of Frauke Petry.

13. Topic 37: Covering debates of AfD and DIE LINKE in the parliament (Deutscher Bundestag).

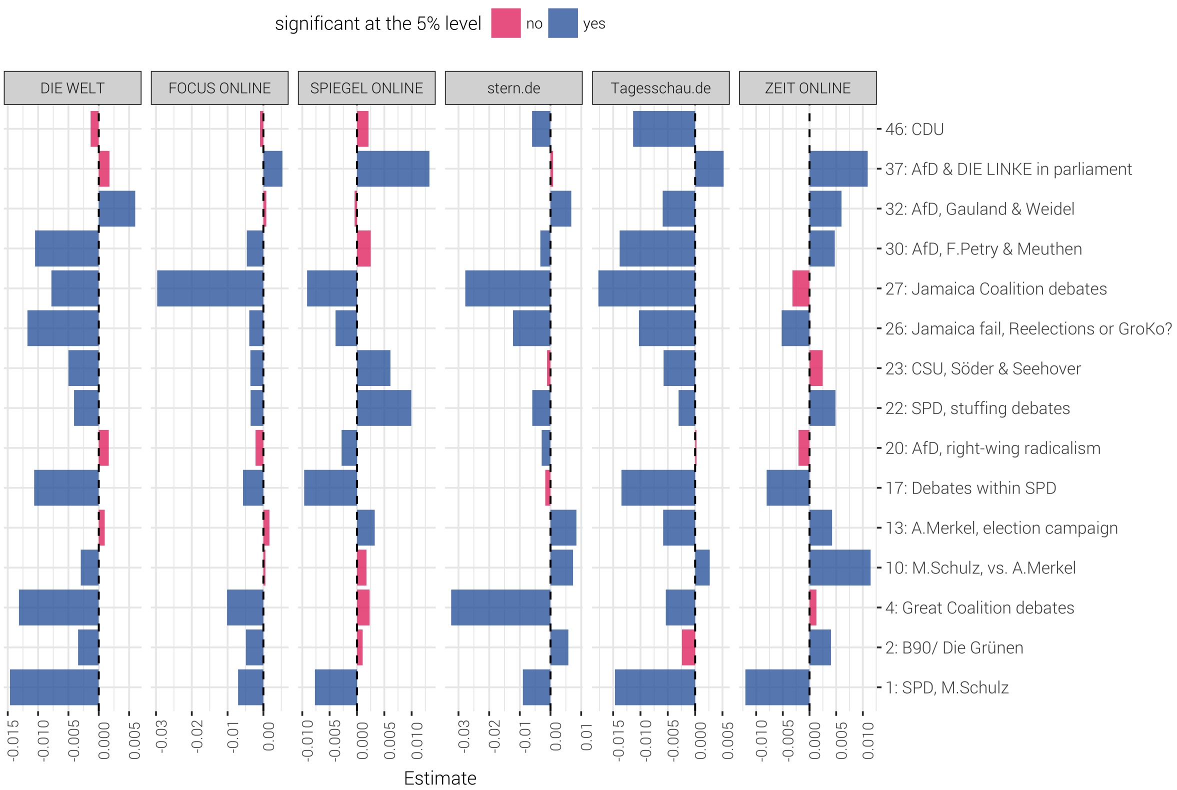
14. Topic 46: Covering issues regarding the CDU/CSU.

To estimate the differences of topic prevalence of the mentioned topics for the different websites, a linear model is calculated for each topic k, where the documents are observations, the dependent variable is the posterior probability of the respective topic (θ_d) and the covariates are dummy variables that are 1 if the document was published by the respective website and 0 otherwise (see equation 5). As the dependent variable is a vector containing proportions that sum up to 1, the QR decomposition of the model is computed prior to the estimation of the coefficients. To incorporate uncertainty in the dependent variable, a set of topic proportions are drawn from the variational posterior (the unnormalized topic proportions) repeated times. Then, the coefficients are computed as the average over all results (M. E. Roberts, B. M. Stewart, and E. M. Airoldi, 2016).

$$\begin{aligned} \theta_d = \beta_0 + \beta_1 x_{\text{FOCUS ONLINE,d}} + \beta_2 x_{\text{SPIEGEL ONLINE,d}} + \\ \beta_3 x_{\text{stern.de,d}} + \beta_4 x_{\text{Tagesschau.de,d}} + \beta_5 x_{\text{DIE WELT,d}} + \\ \beta_6 x_{\text{ZEIT ONLINE,d}} + \epsilon \end{aligned} \tag{5}$$

Figure 7 shows the regression results for the above selected topics (See Appendix A.2 for the result tables). The coefficients indicate the deviation from the base value of Bild.de. Starting from above it becomes apparent that the topic prevalence of topic 46 (regarding the CDU/CSU) is significantly less for Tagesschau.de and Stern.de. The other media do not show any significant difference to Bild.de for this topic. The opposite is true for topic 37: With the exception of Stern.de and DIE WELT, topic prevalence for this topic is significantly higher for all media than for Bild.de. With the following two topics on AfD it is striking that the topic prevalence at Tagesschau.de is significantly lower compared to Bild.de. The topics concerning the Jamaican coalition (topic 27) and the failure (topic 26) seem to be discussed most likely at Bild.de. The case is different for the CSU issue (Topic 23), where SPIEGEL ONLINE has the highest probability. The same applies to the topic related to the personnel debates of the SPD (22). However, Bild.de has the highest topic prevalence for the topic related to votes within the SPD, especially the vote on the grand coalition. The same applies to the topic regarding the SPD in general and Martin Schulz in particular (1). Overall, topics concerning the SPD seem to be more frequent at Bild.de than in the other media. Moreover, the distribution of topics at FOCUS ONLINE seems to be the most similar to that of Bild.de, while the biggest differences exist between Bild.de and Tagesschau.de.

Figure 7: Regression results



5.2 Tone

The sentiment analysis is performed with the documents for which one of the above topics has the highest posterior probability and if this probability is greater than 30%. A dictionary-based method is then applied on the remaining 5611 documents with the aim to measure the tone (or sentiment) of a document. The idea of a sentiment analysis is to determine the attitude of a writer toward the overall tonality of a document. To conduct such an analysis, a lists of words (dictionary) associated with a given emotion, such as negativity is pre-defined by the analyst. The document is then deconstructed into individual words and the frequencies of words contained in a given dictionary are calculated.

Such lexical or “bag-of-words” approaches are widely presented in the finance literature to determine the effect of central banks’ monetary policy communications on asset prices and real variables (Nyman et al. (2018) (Tetlock, 2007), (Tetlock, Saar-Tsechansky, and Macskassy, 2008)). Hansen and McMahon (2016) use a similar approach to measure "the two Ts" (Topic and tone). They explore the effects of FOMC (Federal Open Market Committee) statements on both market and real economic variables. To understand the multi-dimensional information a statement is transmitting, they apply LDA on a corpus of 142 FOMC decision statements split into sentences (topic). They then measure how the central bank is talking about that topic, using a dictionary approach (tone). To calculate their score, they subtract the negative words from the positive words und divide this by the number of total words of the statement. A similar score is used by Nyman et al. (2018), who measure the effect of narratives and sentiment of financial market text-based data on developments in the financial system. They count the number of occurrences of excitement words and anxiety words and then scale these numbers by the total text size as measured by the number of characters.

The present paper uses a dictionary that lists words associated with positive and negative polarity weighted within the interval of $[-1;1]$. SentimentWortschatz[11], is a publicly available German-language resource for sentiment analysis, opinion mining etc. The current version of SentiWS (v1.8b) contains 1,650 positive and 1,818 negative words, which sum up to 15,649 positive and 15,632 negative word forms including their inflections, respectively. The sentiment score for each document d is calculated based on the weighted polarity values for a word, defined on an interval between -1 and 1. The score is then calculated from the sum of the words in a document (which can be assigned to a word from the dictionary) divided by the total number of words in that document:

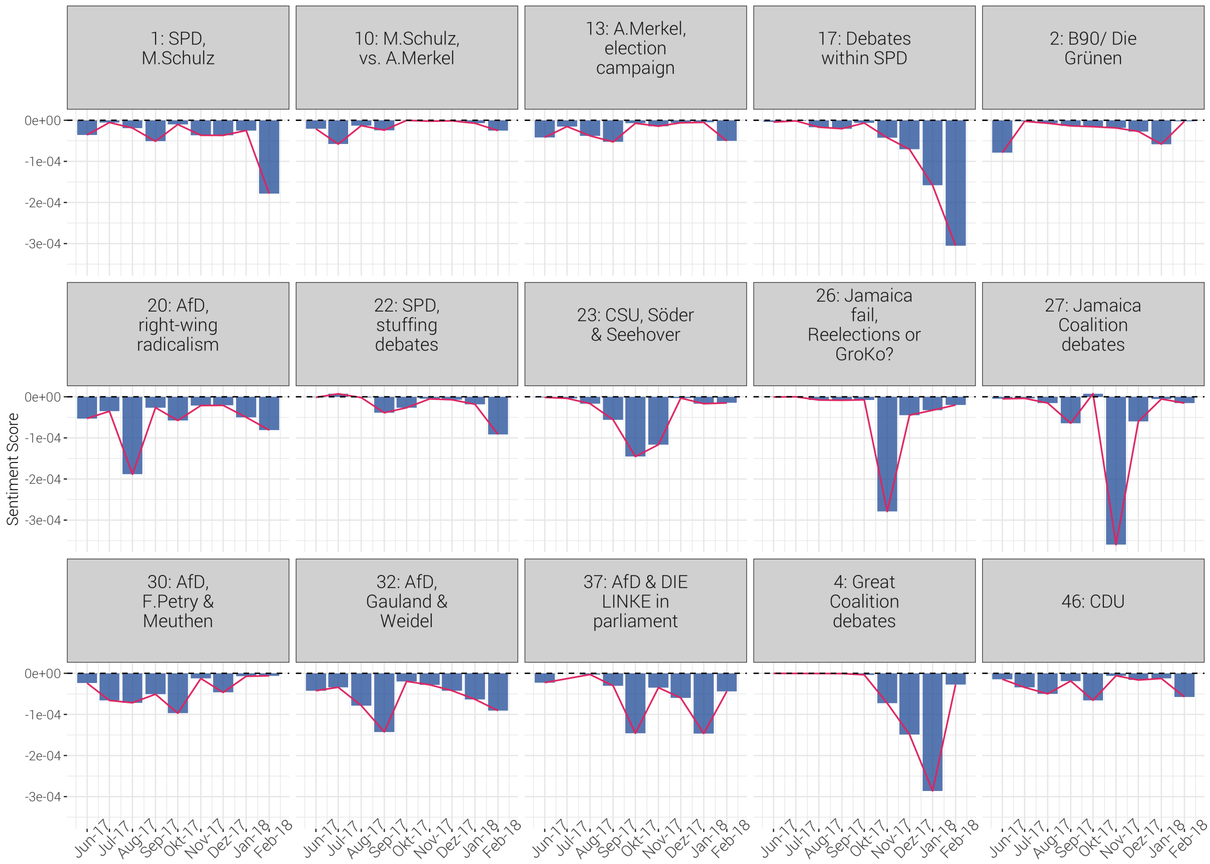
$$\text{SentScore}_d = \frac{|\text{positive polarity score}_d| - |\text{negative polarity score}_d|}{|\text{TotalWords}_d|} \tag{6}$$

Figure 8 shows the results of the analysis for each topic on a monthly basis, aggregated on all newspapers. Each sentiment value is weighted by the relative share of the topic in the overall reporting of that month.

Some conclusions can be drawn from this illustration. First of all, it can be seen that, on average, all topics are discussed almost exclusively negatively. An exception is topic 27 concerning the Jamaica coalition negotiations, which shows a positive sentiment value for a short period of time (October 2017). In the following month (November 2017), after it became clear that there would be no coalition between the CDU/CSU, FDP and Die Grünen, the value of this topic as well as that of topic 26 drops rapidly, but then rises again in February. Concerning the issues that discuss the great coalition between CDU/CSU and SPD, it is evident that the overall tone in which this topic is discussed is generally decreasing from November 2017 to January 2018, but in the following February, the sentiment value of this topic rises. However, the sentiment score of topics that deal with the SPD (1, 17, 22) is diminishing in the course of time, with topic 17 recording the largest decline. For the other parties the process is rather zigzag-like.

[11]SentiWS for short. available here: http://wortschatz.uni-leipzig.de/de/download

Figure 8: Monthly Sentiment Score

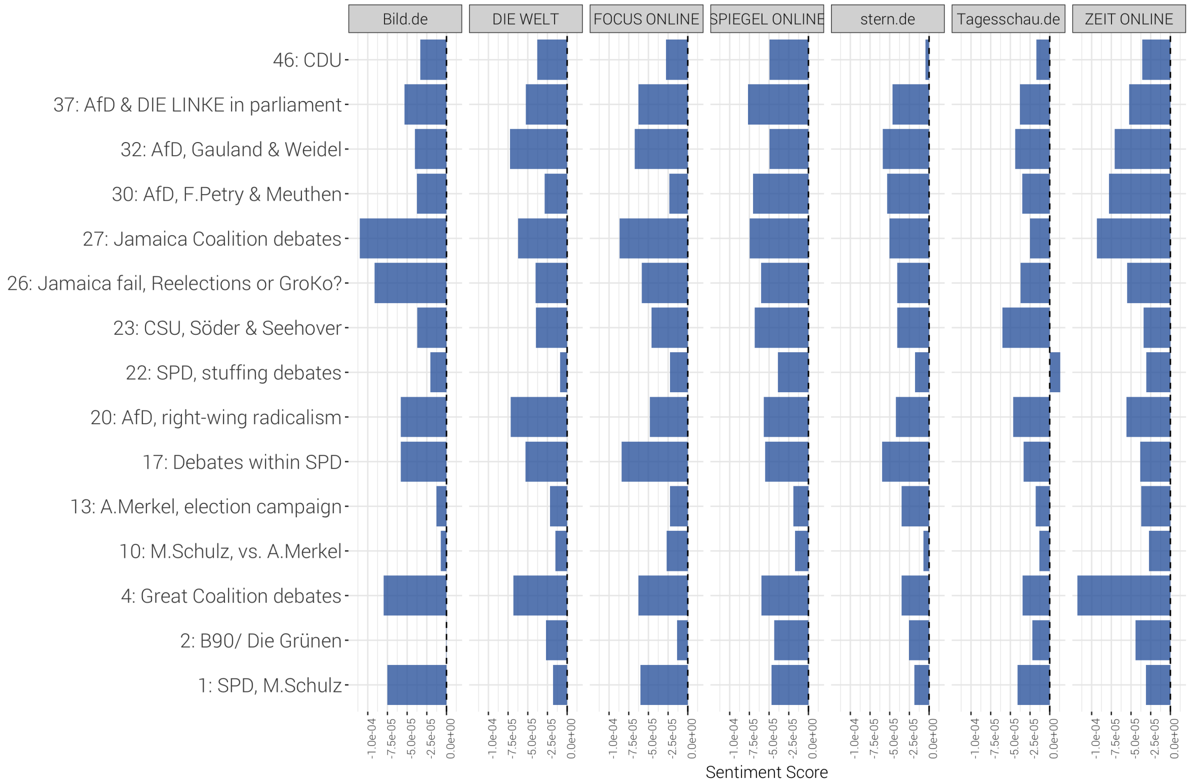


In order to analyze the differences between the news websites, two different figures are considered: The bar plot is used to examine the polarity tendencies of the individual topics for a the respective website (Figure 9) and the radar plot is considered to observe the differences between the websites (Figure 10).

Starting with the bar plot it becomes apparent that all topics are discussed negatively, except topic 23 at Tagesschau.de. At Bild.de, the topics that include the coalition negotiations (26, 27, 4) and the SPD (1, 17) are the most negative. The topics relating to AfD (20,30,32,37) are also discussed more negatively. Looking at the values of DIE WELT, two of the AfD topics have the most negative value (32, 20). Topic 27 concerning the Jamaica Coalition and the Grand Coalition (4) also score relatively negatively. Concerning FOCUS ONLINE, it is mainly topics that relate to the SPD (27, 17, 4, 1) that have a strong negative sentiment value, together with topic 32 and 37 - both related to AfD. Turning to SPIEGEL ONLINE, it is noticeable that the difference in sentiment value between the individual topics is less pronounced. Topics 13 (election campaign of A.Merkel) and 10 (A.Merkel vs M.Schulz) stand out as comparatively less negative. However, these issues are also the least negatively discussed in the other media. Also at stern.de the difference in sentiment value is less significant and overall less negative. The topics regarding CDU/CSU (46) and Martin Schulz (10) score the most positively (or least negatively). Tagesschau.de is the least negative on most topics, or even once positive. However, this does not apply to topic 23 (CSU), where tagesschau.de is most negative in comparison to the other media. As with Bild.de, the issues relating to the coalition negotiations (27 and 4) also come off rather badly with ZEIT ONLINE. However, the issues surrounding AfD

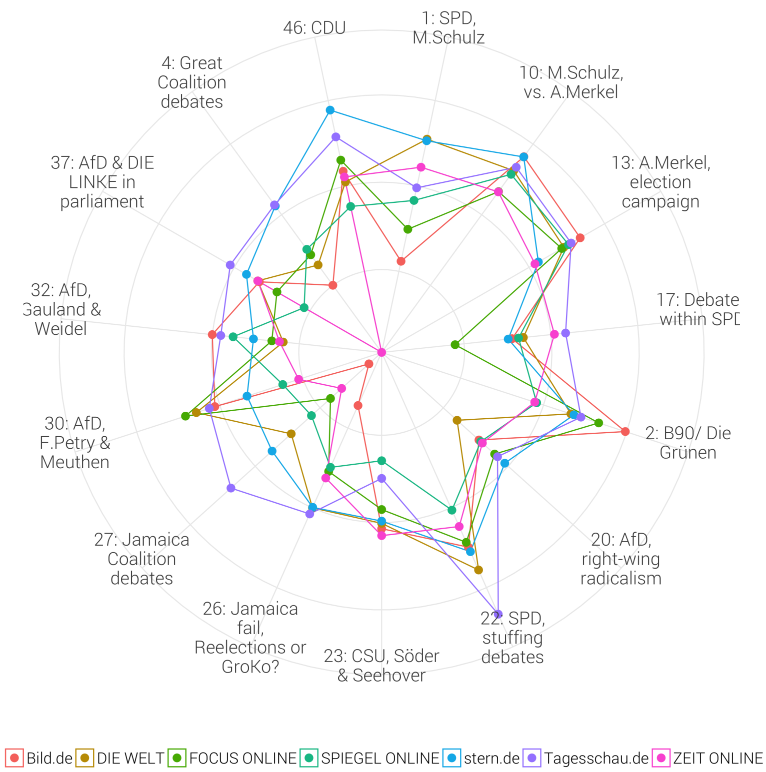
(30, 32, 37 and 20) are even more negative than at Bild.de.

Figure 9: Sentiment Score by news website



A good overview of how differently the topics are discussed by the providers is shown in Figure 10. It becomes evident that the sentiment value of the media differs most notably with regard to topic 27 and topic 4, i.e. the topics on which the coalition negotiations are reported. With regard to the Jamaica coalition, Bild.de reports the most and tagesschau.de the least negative. The reporting of ZEIT ONLINE concerning the grand coalition is the one with the most negative sentiment value and again Tagesschau.de, together with stern.de, the one with the value which is least negative. Furthermore, it becomes evident that the negative sentiment value of FOCUS ONLINE regarding topic 17 is high in relation to the other media. FOCUS ONLINE thus reports comparatively more negatively on the debates within the SPD. This includes in particular the vote on a possible coalition with CDU/CSU/CSU. For topic 1, which also deals with the SPD, the value of FOCUS ONLINE is rather negative, only undercut by Bild.de. Topics related to AfD do not show striking differences.

Figure 10: Radar Plot of Sentiment Scores



After the above figures have been analyzed, the following points can be summarized:

1. The sentiment value of the SPD is decreasing over time, especially regarding debates within the party (topic 17).
2. The topics relating to the coalition talks on Jamaica (26, 27) and the grand coalition (4) are discussed rather critically, but they also show the greatest differences between the media.
3. In contrast, the tonality of the topics in relation to the AfD shows rather small differences.
4. Overall, the sentiment value at Tagesschau.de is the least negative and only shows a comparatively strong negative value at topic 23, concerning the CSU.

5.3 News sentiment and poll data

This section seeks to examine the association between sentiment reflected in online news content and phone survey poll results in Germany. Specifically, it aims to find the extent to which online sentiment and phone survey results correlate given a number of lags. I use the data from the "Sonntagsumfrage" (Sunday survey) from infratest dimap.[12] The institution regularly asks at least 1000 German citizens the question: "Which party would you choose if federal elections take place next Sunday?" The survey thus measures the current election tendencies and therefore reflects an intermediate state in the opinion-forming process of the electoral population.

[12] https://www.infratest-dimap.de/umfragen-analysen/bundesweit/sonntagsfrage/

Much of the research on online content and political trends have focused on traditional weblogs and social media websites, such as Twitter, Facebook, MySpace, and YouTube. These studies have shown that social media is used to spread political opinions and that these considerations reflect the political landscape of the offline world. Tumasjan et al. (2010) investigate Tweets between August 13th and September 19th, 2009, prior to the German national elections to examine whether Twitter messages reflect the current offline political sentiment and whether it can be used to predict the popularity of parties or coalitions in the real world. With regard to the later question, they compare the share of attention the political parties receive on Twitter with the election result to examine whether the activity on Twitter can serve as a predictor of the election outcome. They found that the number of tweets reflects the election result and even comes close to traditional election polls.

Fu and Chan (2013) use a corpus of online posts from discussion forums and blogs to examine the extent to which online sentiment reflected in social media content can predict phone survey results in Hong Kong. They build a sentiment classifier conducting a support vector machine analysis on a training set of 2,000 manually labeled posts. In order to evaluate the temporal relationship between the time series of the online sentiment score and the results of the telephone survey, a cross correlation analysis was conducted, using the Box and Jenkins autoregressive integrated moving average (ARIMA) method (Box, Jenkins, and Reinsel, 2008). Estimating the cross-correlation functions of the residuals, they find that online sentiment scores can lead phone survey results by about 8–15 days.

In a more recent conference paper, Padmaja, Fatima, and Bandu (2014) identify the scope of negation in news articles for two political parties in India (BJP and UPA) to analyze how the choice of certain words used in these texts influence the sentiments of public in polls. Comparing three different sentiment analysis methods (two machine learning and one dictionary method), they observe that the choice of certain words used in political text was influencing the sentiments in favor of BJP. They conclude that this sentiment bias might be one of the causes for the election results in 2014.

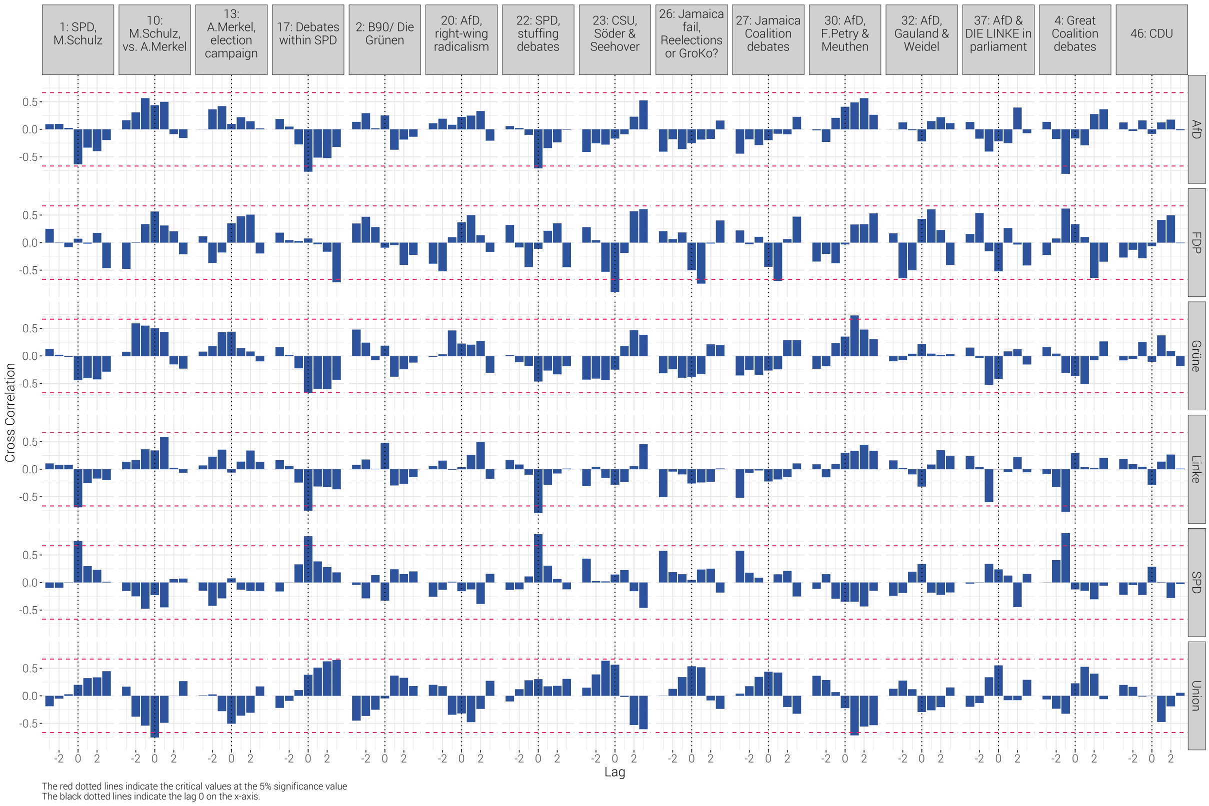
In the present paper, the relationship between monthly average of both the sentiment value of individual topics (x_t) and the survey value of the parties (y_t) is estimated using the cross correlation function (CCF). Thus, the CCF between x_{t+h} and y_t for $h \pm 1, h \pm 2, h \pm 3$ is computed. A negative value for h is a correlation between the topic sentiment value at a time before t and the survey value at time t. The correlation value for $h = 0$ indicates the contemporary correlation between the two time series. Based on the coefficients of the cross correlation estimation shown in Figure 11, the significant correlations between topic sentiment and survey value are evaluated for each party.[13] It is important to note that no causal relationships are described below, but that only the correlation between the two time series is described.

The survey results of the AfD correlate negatively with topics relating to the SPD (17, 22) at lag 0. Thus, if the SPD was more negatively reported, the poll value of the AfD increased in the same month (and vice versa). Another significant negative correlation exists between the reporting on the GroKo (4) and the survey value of AfD at lag -1 (x_{t-1}). So if the GroKo was more negatively reported in one month, the survey value of the AfD increased in the following month (and vice versa). For the FDP, too, only negative correlation coefficients can be detected, with the strongest negative correlation existing for the topic relating to the CSU (23). If the CSU got off worse in the online news, the poll value of the FDP have gone up. Another interesting observation is that the FDP's poll results correlate negatively with issues relating to the Jamaica coalition at lag 1 (x_{t+1}). So if the poll results for the FDP rose in one month, the following month

[13]The value of the cross correlation coefficients for lag 0 can be found in the Appendix (Table 2)

the FDP was reported more negatively. The Green Party survey results show no negative correlation with any of the topics, except topic 30 at lag 1. It is striking that there seems to be a strong negative correlation between the SPD topics (1, 17, 22) and the poll results of the left party (DIE LINKE). This means that the poll value of the left party has climbed if the topics related to the SPD were discussed more negatively. Same applies to the reporting on the GroKo (30) for lag -1. Conversely, the SPD's survey results correlate strongly positively with these topics, and also with topic 30 with a delay of one month. For the CDU/CSU, too, only significant negative correlations are discernible: the survey results correlate negatively with the topic of the Schulz v Merkel debate (10) and negatively with topic 30 with a delay of one month (x_{t+1}).

Figure 11: Cross-Correlation



After the above figure have been analyzed, the following points can be summarized:

1. Only the survey results of the SPD correlate positively with the emotional value of the topics. There seems to be a strong correlation between the way topics concerning the SPD are discussed in the online news and the poll results.

2. The poll results of the Left Party, on the other hand, seem to correlate negatively with the reporting on the SPD.

3. Similar tendencies can also be seen with regard to the AfD, since here too the survey results correlate significantly negatively with the topics about the SPD and the grand coalition.

Summarizing the analyses from this and the previous section, it can be observed that the positive correlation between the emotional value of the reporting and the survey value of a party is particularly large if the reporting is conspicuously negative.

6 Conclusion

The purpose of this paper was to examine (1) whether the political reporting of different media differs in terms of topic frequency and topic tonality and (2) whether this reporting correlates with the opinion-forming process of the voters. Regarding (1) the analysis revealed that there are differences between the media considered, both in terms of topic prevalence and the way in which these topics are discussed. Although overall all topics are discussed negatively, there are still differences, especially regarding the coalition negotiations. The smallest differences can be found for topic concerning the AfD. With regard to (2), the analysis has shown that the tonality of topics discussed by the SPD shows a strong positive correlation to current survey results. Overall, there seems to be a link between reporting on political issues and electoral preferences. Further research should focus on the exact causal relationships between these two concepts.

References

Baturo, Alexander, Niheer Dasandi, and Slava J. Mikhaylov (2017). "What Drives the International Development Agenda? An NLP Analysis of the United Nations General Debate 1970-2016". In: *arXiv:1708.05873 [cs]*. arXiv: 1708.05873. URL: http://arxiv.org/abs/1708.05873.

Bholat, David M. et al. (2015). "Text Mining for Central Banks". In: *SSRN Electronic Journal*. URL: http://www.academia.edu/13430482/Text_mining_for_central_banks (visited on 11/06/2017).

Blei, David M. (2012). "Probabilistic Topic Models". In: *Commun. ACM* 55.4, pp. 77–84. URL: http://doi.acm.org/10.1145/2133806.2133826.

Blei, David M., Andrew Y Ng, and Michael I Jordan (2003). "Latent dirichlet allocation". In: *Journal of machine Learning research* 3, pp. 993–1022.

Box, George E. P., Gwilym M. Jenkins, and Gregory C. Reinsel (2008). *Time series analysis: forecasting and control*. 4th ed. Wiley series in probability and statistics. Hoboken, NJ: Wiley.

Braun, Michael and Jon McAuliffe (2010). "Variational inference for large-scale models of discrete choice". In: *Journal of the American Statistical Association* 105.489. arXiv: 0712.2526, pp. 324–335. URL: http://arxiv.org/abs/0712.2526 (visited on 01/19/2018).

Cagé, Julia, Nicolas Hervé, and Marie-Luce Viaud (2017). *L'information à tout prix*. fr. Éditions de l'INA.

Ellman, Matthew and Fabrizio Germano (2009). "What do the Papers Sell? A Model of Advertising and Media Bias*". en. In: *The Economic Journal* 119.537, pp. 680–704. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0297.2009.02218.x/abstract.

Erosheva, Elena, Stephen Fienberg, and John Lafferty (2004). "Mixed-membership models of scientific publications". en. In: *Proceedings of the National Academy of Sciences* 101.suppl 1, pp. 5220–5227. URL: http://www.pnas.org/content/101/suppl_1/5220 (visited on 10/12/2017).

Farrell, Justin (2016). "Corporate funding and ideological polarization about climate change". en. In: *Proceedings of the National Academy of Sciences* 113.1, pp. 92–97. URL: http://www.pnas.org/content/113/1/92 (visited on 11/09/2017).

Fu, King-wa and Chee-hon Chan (2013). "Analyzing Online Sentiment to Predict Telephone Poll Results". en. In: *Cyberpsychology, Behavior, and Social Networking* 16.9, pp. 702–707. URL: http://online.liebertpub.com/doi/abs/10.1089/cyber.2012.0375 (visited on 03/19/2018).

Gabszewicz, Jean J., Dider Laussel, and Nathalie Sonnac (2001). "Press advertising and the ascent of the 'Pensée Unique'". In: *European Economic Review*. 15th Annual Congress of the European Economic Association 45.4, pp. 641–651. URL: http://www.sciencedirect.com/science/article/pii/S0014292101001398.

Gentzkow, Matthew, Bryan T. Kelly, and Matt Taddy (2017). *Text as Data*. Working Paper 23276. National Bureau of Economic Research. URL: http://www.nber.org/papers/w23276.

Griffiths, Thomas L. and Mark Steyvers (2002). "A probabilistic approach to semantic representation". In: *Proceedings of the Annual Meeting of the Cognitive Science Society* 24.24. URL: https://escholarship.org/uc/item/44x9v7m7 (visited on 11/16/2017).

Griffiths, Thomas L. and Mark Steyvers (2004). “Finding scientific topics”. en. In: *Proceedings of the National Academy of Sciences* 101.suppl 1, pp. 5228–5235. URL: http://www.pnas.org/content/101/suppl_1/5228 (visited on 10/12/2017).

Grimmer, Justin and Brandon Stewart (2013). “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts”. In: *Political Analysis* 21, pp. 267–297.

Hansen, Stephen and Michael McMahon (2016). “Shocking language: Understanding the macroeconomic effects of central bank communication”. In: *Journal of International Economics*. 38th Annual NBER International Seminar on Macroeconomics 99, S114–S133. URL: http://www.sciencedirect.com/science/article/pii/S0022199615001828 (visited on 03/07/2018).

Hofmann, Thomas (1999). “Probabilistic Latent Semantic Indexing”. In: *Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR ’99. New York, NY, USA: ACM, pp. 50–57. URL: http://doi.acm.org/10.1145/312624.312649.

Verbrauchs- und Medienanalyse - VuMA 2018 (2017). de. Tech. rep. URL: https://de.statista.com/statistik/daten/studie/183128/umfrage/nachrichten-und-informationen---internetnutzung/ (visited on 03/12/2018).

Mimno, David et al. (2011). “Optimizing Semantic Coherence in Topic Models”. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. EMNLP ’11. Stroudsburg, PA, USA: Association for Computational Linguistics, pp. 262–272. URL: http://dl.acm.org/citation.cfm?id=2145432.2145462.

Mishler, Alan et al. (2015). “Using Structural Topic Modeling to Detect Events and Cluster Twitter Users in the Ukrainian Crisis”. en. In: *HCI International 2015 - Posters’ Extended Abstracts*. Communications in Computer and Information Science. Springer, Cham, pp. 639–644. URL: https://link.springer.com/chapter/10.1007/978-3-319-21380-4_108 (visited on 10/12/2017).

Mueller, Hannes Felix and Christopher Rauh (2016). *Reading between the Lines: Prediction of Political Violence Using Newspaper Text*. SSRN Scholarly Paper ID 2843535. Rochester, NY: Social Science Research Network. URL: https://papers.ssrn.com/abstract=2843535 (visited on 11/09/2017).

Nyman, Rickard et al. (2018). “News and narratives in financial systems: exploiting big data for systemic risk assessment | Bank of England”. In: *Bank of England Working Paper* 704. URL: https://www.bankofengland.co.uk/working-paper/2018/news-and-narratives-in-financial-systems (visited on 02/21/2018).

Padmaja, S., Prof S. Sameen Fatima, and Sasidhar Bandu (2014). “Evaluating Sentiment Analysis Methods and Identifying Scope of Negation in Newspaper Articles”. en. In: *International Journal of Advanced Research in Artificial Intelligence (IJARAI)* 3.11. URL: http://thesai.org/Publications/ViewPaper?Volume=3&Issue=11&Code=IJARAI&SerialNo=1 (visited on 03/19/2018).

Paul, Michael (2017). “Cross-Collection Topic Models: Automatically Comparing and Contrasting Text”. In:

Pritchard, J. K., M. Stephens, and P. Donnelly (2000). “Inference of population structure using multilocus genotype data”. eng. In: *Genetics* 155.2, pp. 945–959.

Roberts, Margaret E., Brandon M. Stewart, and Edoardo M. Airoldi (2016). “A Model of Text for Experimentation in the Social Sciences”. In: *Journal of the American Statistical Association* 111.515, pp. 988–1003. URL: http://dx.doi.org/10.1080/01621459.2016.1141684.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, et al. (2014). “Structural Topic Models for Open-Ended Survey Responses”. en. In: *American Journal of Political*

Science 58.4, pp. 1064–1082. URL: http://onlinelibrary.wiley.com/doi/10.1111/ajps.12103/abstract.

Roberts, Margaret, Brandon Stewart, and Dustin Tingley (2016a). "Navigating the Local Modes of Big Data: The Case of Topic Models." In: *Computational Social Science: Discovery and Prediction*. New York: Cambridge University Press.

— (2016b). "stm: R Package for Structural Topic Models". In: *Journal of Statistical Software* forthcoming.

Roberts, Margaret, Brandon Stewart, Dustin Tingley, and Edoardo Airoldi (2013). "The Structural Topic Model and Applied Social Science". In: *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation.*

Steiner, Peter O. (1952). "Program Patterns and Preferences, and the Workability of Competition in Radio Broadcasting". In: *The Quarterly Journal of Economics* 66.2, pp. 194–223. URL: http://www.jstor.org/stable/1882942.

Tetlock, Paul C. (2007). "Giving Content to Investor Sentiment: The Role of Media in the Stock Market". en. In: *The Journal of Finance* 62.3, pp. 1139–1168. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2007.01232.x/abstract.

Tetlock, Paul C., Maytal Saar-Tsechansky, and Sofus Macskassy (2008). "More Than Words: Quantifying Language to Measure Firms' Fundamentals". en. In: *The Journal of Finance* 63.3, pp. 1437–1467. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.2008.01362.x/abstract (visited on 03/07/2018).

Tumasjan, Andranik et al. (2010). "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment". en. In: *Proceedings of the Fourth International Conference on Weblogs and Social Media*. Washington. URL: https://www.researchgate.net/publication/215776042_Predicting_Elections_with_Twitter_What_140_Characters_Reveal_about_Political_Sentiment (visited on 03/17/2018).

Wiedmann, Gregor (2016). *Text Mining for Qualitative Data Analysis in the Social Sciences*. 1st ed. Wiesbaden: VS Verlag für Sozialwissenschaften. URL: //www.springer.com/de/book/9783658153083 (visited on 11/26/2017).

A Appendices

A.1 Most frequent words

1: SPD

Bild.de : schulz spd martin gabriel partei merkel chef kanzlerkandidat wahlkampf bundestagswahl wahl sigmar angela politik andrea parteitag genossen scholz berlin kanzlerin

focus.de : schulz spd martin partei wahlkampf merkel kanzlerkandidat schröder politik bundestagswahl wahl chef scholz tv angela kanzler sozialdemokraten chanc parteitag wochen

spiegel.de : schulz spd martin kanzlerkandidat duell wahlkampf partei tv bundestagswahl merkel politik wahl sozialdemokraten umfragen chef kanzlerkandidaten wochen herausforder auftritt kanzler

stern.de : schulz spd martin duell tv merkel kanzlerkandidat wahlkampf angela septemb bundestagswahl berlin partei herausforder wahl kanzlerin strunz politik zdf sonntag

tagesschau.de : schulz spd partei martin chef parteitag parteichef merkel scholz kanzlerkandidat genossen opposit groko juso müsse bundestagswahl brandt wahl wahlkampf jahr

welt.de : schulz spd martin gabriel kanzlerkandidat wahlkampf partei merkel welt bundestagswahl wahl sozialdemokraten berlin außenminist duell chef heil wochen schröder kanzlerkandidaten

zeit.de : schulz spd martin partei wahlkampf kanzlerkandidat genossen merkel berlin marcu chef glahn scholz bundestagswahl sozialdemokraten parteitag angela parteichef wahl gabriel

2: B90/ Die Grüne & DIE LINKE

Bild.de : grünen özdemir göring eckardt partei habeck wagenknecht cem wahlkampf parteitag lilli linken katrin grüne link chef wahl blome politik talk

focus.de : grünen özdemir grüne partei habeck göring eckardt cem parteitag linken robert flügel berlin realo katrin baerbock schleswig wahl trittin parteichef

spiegel.de : grünen özdemir habeck göring eckardt partei grüne parteitag katrin cem baerbock robert linken link flügel politik hofreit realo wahl koalit

stern.de : grünen özdemir grüne göring link eckardt partei cem parteitag linken rot habeck bundestagswahl wagenknecht katrin berlin koalit fdp grün spitzenkandidatin

tagesschau.de : grünen habeck partei özdemir top göring grüne wagenknecht eckardt parteitag baerbock linken flop cem linkspartei peter katrin parteichef robert rot

welt.de : grünen özdemir habeck partei linken göring eckardt rot grüne robert cem link parteitag grün katrin peter welt baerbock schleswig wagenknecht

zeit.de : grünen özdemir partei linken link göring habeck eckardt parteitag grüne cem hofreit katrin realo baerbock rot politik wagenknecht koalit peter

4: Great Coalition debates

Bild.de : spd union groko cdu csu koalit schulz koalitionsverhandlungen chef parteitag gespräch verhandlungen große großen martin sondierungen regierung berlin ergebni partei

focus.de : spd union groko cdu koalit csu koalitionsverhandlungen schulz parteitag große verhandlungen chef großen sondierungen koalitionsvertrag gespräch sozialdemokraten partei martin stegner

spiegel.de : spd union cdu csu koalit groko koalitionsverhandlungen schulz parteitag verhandlungen politik großen sondierungen chef sozialdemokraten berlin große gespräch martin koalitionsvertrag

stern.de : spd union cdu csu verhandlungen koalit koalitionsverhandlungen schulz sondierungen groko berlin großen martin gespräch chef press merkel parteien große einigung

tagesschau.de : spd union cdu csu koalitionsverhandlungen koalit groko schulz verhandlungen chef sondierungen parteitag gespräch großen große sozialdemokraten martin koalitionsvertrag partei parteien

welt.de : spd union koalit cdu csu groko koalitionsverhandlungen große schulz welt großen gespräch sondierungen verhandlungen chef parteitag martin sozialdemokraten berlin bürgerversicherung

zeit.de : spd union koalit cdu csu große koalitionsverhandlungen schulz großen parteitag verhandlungen chef koalitionsvertrag sondierungen gespräch partei juso sozialdemokraten martin kompromiss

10: Merkel vs. Schulz

Bild.de : schulz spd merkel martin duell kanzlerin deutschland tv thema flüchtling kanzlerkandidat angela rent wahlkampf zitat cdu eu cichowicz pflege union

focus.de : schulz spd thema duell merkel tv martin wahlkampf deutschland kanzlerkandidat kanzlerin herausforder flüchtling angriff frage union menschen cdu rent themen

spiegel.de : schulz spd martin merkel eu kanzlerkandidat deutschland europa duell thema italien kanzlerin tv flüchtlingskris maut flüchtling angela wahlkampf politik union

stern.de : schulz spd merkel martin kanzlerkandidat deutschland flüchtlingskris italien kanzlerin europa bildung union thema eu angela flüchtling cdu bundestagswahl berlin geld

tagesschau.de : schulz duell tv spd merkel thema siemen martin deutschland rent kanzlerkandidat herausforder maut menschen flüchtling chef konzern kanzlerin kaeser union

welt.de : schulz spd merkel martin duell deutschland thema tv flüchtling kanzlerkandidat cdu land welt stimmt kanzlerin türkei beitrittsverhandlungen italien angela union

zeit.de : schulz spd duell merkel martin tv kanzlerkandidat deutschland angela kanzlerin thema wahlkampf flüchtling eu themen union rent cdu bundestagswahl europa

13: A.Merkel

Bild.de : merkel kanzlerin angela cdu bundeskanzlerin union wahlkampf deutschland rede auftritt jahr politik bundestagswahl kanzleramt frage chefin tv zdf kurz spd

focus.de : merkel kanzlerin angela cdu bundeskanzlerin union deutschland thema frage wahlkampf menschen wahl jahr bundestagswahl abstimmung kurz jahren müsse politik spd

spiegel.de : merkel kanzlerin angela cdu bundeskanzlerin wahlkampf union chefin auftritt deutschland sebastian frage kanzleramt jahr zdf jahren thema berlin spd kurz

stern.de : merkel angela kanzlerin cdu bundeskanzlerin union wahlkampf deutschland abstimmung bundestagswahl menschen thema frage jahr septemb berlin spd wahlkampfauftritt rede politik

tagesschau.de : merkel kanzlerin angela cdu bundeskanzlerin thema chefin wahlkampf union abstimmung jahr deutschland kanzleramt brigitt forderung frage paar menschen stand bundestagswahl

welt.de : merkel angela kanzlerin cdu bundeskanzlerin deutschland union zdf duell wahlkampf tv welt regierung jahr frage sendung menschen peter jahren chefin

zeit.de : merkel angela kanzlerin cdu bundeskanzlerin wahlkampf union deutschland thema kritik frage menschen jahr jahren bundestagswahl spd kurz flüchtlingspolitik politik könne

ntra-party vote (SPD) about the coalition w C

Bild.de : spd schulz nahl groko gabriel partei chef koalit juso kühnert mitglied martin außenminist kevin große andrea koalitionsvertrag sigmar parteitag union

focus.de : spd gabriel schulz außenminist partei sigmar juso koalit groko kühnert chef großen große union martin mitglied amt nahl parteitag sozialdemokraten

spiegel.de : spd gabriel schulz partei mitglied außenminist koalit sigmar politik groko chef juso martin koalitionsvertrag sozialdemokraten große union großen parteitag regierung

stern.de : spd schulz gabriel koalit partei groko martin union chef große juso parteitag sigmar großen außenminist koalitionsvertrag koalitionsverhandlungen kühnert sozialdemokraten mitglied

tagesschau.de : spd gabriel schulz partei koalit außenminist groko koalitionsvertrag chef großen martin große sigmar europa union nahl sozialdemokraten mitglied parteichef politik

welt.de : spd schulz gabriel partei welt koalit nahl mitglied groko juso koalitionsvertrag sigmar außenminist martin kühnert andrea chef großen große cdu

zeit.de : spd gabriel partei schulz sigmar koalit außenminist klingbeil martin union große großen parteitag juso sozialdemokrati mitglied chef bundestagswahl politik sozialdemokraten

20: AfD, Right-wing extremist tendencies

Bild.de : afd storch twitter facebook account plakat tweet partei politik beatrix maier schrieb amann wahlbetrug post weißig jahr wahlplak deutschland show

focus.de : afd tweet twitter facebook partei politik vw aktion schrieb storch montag gruppen jahr arpp vorwürf rechten erklärt volkswagen terror account

spiegel.de : afd vw hampel politik partei staatsanwaltschaft storch arpp rede volkswagen regierungserklärung konzern berlin vorwürf tweet affär strafanzeig landesvorstand bundesvorstand jahr

stern.de : afd vw facebook partei gruppen post hampel nrw storch antifa land berlin septemb landesverband niedersächsischen beatrix wahlkampf tweet vorwürf spd

tagesschau.de : afd account vw ndr twitter hampel fake arpp roth zitat partei suhren falsch ministerpräsid tweet informationen bystron becker regierungserklärung pörksen

welt.de : afd hampel pegida politik welt partei landesverband plakat landesvorstand parteitag niedersachsen niedersächsischen storch mecklenburg vorpommern bundesvorstand npd brief landeslist list

zeit.de : afd jahr beruf funktion früher mitgli partei unbekannt arpp orientierung nationalkonservativ berlin sachsen landesverband identitären höcke nordrhein cdu björn bayern

23: CSU, Söder vs. Seehover, refugee ca

Bild.de : csu seehof söder horst bayern cdu chef union herrmann marku parteichef partei berlin ministerpräsid obergrenz bundestagswahl parteitag landtagswahl bayerisch jahr

focus.de : csu seehof söder horst obergrenz union cdu partei bayern chef berlin ministerpräsid bundestagswahl marku parteichef parteitag jahr landtagswahl bayerisch montag

spiegel.de : csu seehof söder horst cdu obergrenz union bayern partei berlin chef bundestagswahl ministerpräsid marku politik jahr bayerisch parteichef parteitag dobrindt

stern.de : csu seehof horst union cdu bundestagswahl obergrenz söder berlin chef bayern jamaika dobrindt partei ministerpräsid koalit grünen marku bayerisch alexand

tagesschau.de : csu seehof söder obergrenz horst union partei cdu berlin bundestagswahl chef bayern ministerpräsid marku bayerisch parteichef herrmann dobrindt offen parteitag

welt.de : csu seehof söder bayern horst cdu union obergrenz herrmann partei chef marku berlin bundestagswahl ministerpräsid bayerisch welt landtagswahl jahr parteichef

zeit.de : csu seehof obergrenz söder horst cdu union bayern chef partei ministerpräsid bundestagswahl bayerisch herrmann dobrindt berlin marku parteichef jahr landtagswahl

Reelections vs. Great Coalition after Jamaic

Bild.de : spd merkel jamaika neuwahlen koalit groko minderheitsregierung regierung bundestag schulz kanzlerin angela union scheitern parteien große chef cdu gespräch steinmeier
focus.de : koalit spd neuwahlen jamaika merkel große minderheitsregierung steinmeier union regierung groko schulz scheitern bundespräsid großen verantwortung kanzlerin parteien partei frank
spiegel.de : koalit spd jamaika merkel neuwahlen minderheitsregierung große gespräch steinmeier union regierung schulz scheitern sozialdemokraten bundespräsid großen kanzlerin angela cdu regierungsbildung
stern.de : spd koalit merkel minderheitsregierung jamaika neuwahlen steinmeier große regierungsbildung union regierung großen schulz gespräch angela bundespräsid cdu groko kanzlerin chef
tagesschau.de : koalit spd jamaika merkel minderheitsregierung neuwahlen gespräch große steinmeier regierungsbildung großen union scheitern bundespräsid schulz regierung parteien chef partei gesprächen
welt.de : spd koalit jamaika neuwahlen merkel minderheitsregierung scheitern große gespräch steinmeier regierung union schulz sondierungen neuwahl angela chef großen cdu bundespräsid
zeit.de : koalit spd merkel minderheitsregierung jamaika große scheitern regierung neuwahlen union gespräch großen schulz neuwahl regierungsbildung partei steinmeier opposit cdu angela

27: Jamaica caolition talks

Bild.de : fdp grünen jamaika lindner csu cdu union grüne chef koalit parteien christian verhandlungen kubicki sondierungen gespräch trittin liberalen thema dobrindt
focus.de : fdp jamaika grünen lindner csu union cdu christian chef parteien grüne kubicki sondierungen verhandlungen liberalen koalit gespräch partei bündni deutschland
spiegel.de : fdp grünen lindner jamaika csu cdu union sondierungen grüne chef parteien christian politik koalit kubicki gespräch liberalen trittin partei thema
stern.de : fdp jamaika grünen lindner csu cdu sondierungen scheitern koalit union parteien gespräch christian grüne chef deutschland neuwahlen berlin abbruch partei
tagesschau.de : fdp grünen jamaika csu lindner union cdu parteien sondierungen grüne koalit chef verhandlungen gespräch themen christian partei kubicki müsse thema
welt.de : fdp grünen jamaika lindner csu cdu welt sondierungen union grüne christian gespräch parteien kubicki chef koalit verhandlungen trittin thema liberalen
zeit.de : fdp grünen jamaika lindner csu union cdu grüne parteien koalit christian sondierungen chef gespräch kubicki sondierungsgespräch liberalen verhandlungen partei thema

30: F.Petri, AfD

Bild.de : afd petri partei meuthen frauk gauland fraktion weidel höcke bundestag bundesvorstand abgeordnet jörg parteitag co pazderski sachsen landtag alexand pretzel

focus.de : afd petri partei frauk fraktion meuthen höcke landtag immunität sachsen gauland bundestag pretzel austritt flügel bundestagswahl jörg parteitag björn verlassen

spiegel.de : afd petri partei frauk meuthen fraktion gauland pretzel bundestag sachsen landtag politik bundestagswahl höcke weidel parteitag poggenburg sächsischen bundestagsfrakt fraktionschef

stern.de : afd petri frauk partei fraktion gauland meuthen bundestag pretzel weidel bundestagswahl landtag höcke alexand dresden austritt vorsitzend chefin jörg alic

tagesschau.de : afd petri partei landtag frauk sachsen fraktion meuthen gauland bundestag höcke staatsanwaltschaft pegida dresden weidel sächsischen poggenburg angekündigt alexand antrag

welt.de : afd petri partei bundestagswahl bundestag berlin cdu frauk wahl fraktion meuthen ergebni gauland direktmandat parlament sachsen deutschland kraft politik alexand

zeit.de : afd petri partei frauk fraktion gauland meuthen landtag bundestag parteitag höcke sachsen weidel septemb maier abgeordnet bundestagswahl luck wahl immunität

32: Gauland, Weidel, AfD

Bild.de : afd gauland weidel alic partei alexand deutschland spitzenkandidatin politik sayn özoguz pazderski spitzenkandidat zeitung land spd wittgenstein demokrati bundestag zdf

focus.de : afd weidel gauland partei politik alexand alic deutschland sendung zdf moderatorin islam studio illner parteien zeitung spitzenkandidatin bundestag talkshow deutsch

spiegel.de : afd gauland weidel partei özoguz politik alexand deutsch alic deutschland rechtspopulisten stiftung land rede höcke entsorgen aydan berlin spitzenkandidat rechten

stern.de : afd gauland weidel partei alic alexand deutschland özoguz höcke deutschen spitzenkandidat deutsch septemb spitzenkandidatin politikerin entsorgen politik zdf äußerungen spd

tagesschau.de : afd gauland weidel partei stiftung meuthen alic deutschland alexand parteitag flügel höcke politik facebook kandidaten pazderski gniffk menschen stiftungen bundestag

welt.de : afd gauland weidel partei alic welt höcke alexand deutschland politik özoguz meuthen parteitag bundestag stiftung wagner pazderski björn deutsch rechten

zeit.de : afd gauland weidel partei höcke alexand deutschland özoğuz alic altern politik äußerungen deutschen nazi stiftung pazderski entsorgen deutsch spitzenkandidatin meuthen

37: AfD in parliament

Bild.de : afd bundestag fraktion abgeordneten wagenknecht abgeordnet glaser fraktionen link parlament spd fdp partei linken politik wahl sitzung diäten gewählt cdu

focus.de : afd bundestag fraktion abgeordneten abgeordnet linken link schäubl wagenknecht sitzung partei parlament bartsch spd wahl politik fraktionen glaser parteien antrag

spiegel.de : bundestag afd link wagenknecht linken fraktion partei politik abgeordnet parlament abgeordneten kip glaser spd fraktionen sahra sitzung fraktionschef wahl fdp

stern.de : bundestag afd fraktion abgeordneten parlament abgeordnet fraktionschef link spd fdp linken fraktionen wagenknecht sitzung politik schäubl cdu bartsch wahl stimmen

tagesschau.de : bundestag afd fraktion abgeordneten parlament sitzung abgeordnet fraktionen glaser linkspartei partei wahl gewählt kandidaten stimmen wagenknecht fdp spd amt schäubl

welt.de : afd bundestag fraktion abgeordneten abgeordnet parlament link fraktionen welt fdp spd partei antrag cdu gewählt wahl politik sitzung stimmen schäubl

zeit.de : afd bundestag abgeordneten abgeordnet fraktion link linken partei glaser spd politik fdp sitzung fraktionen wahl schäubl parlamentarisch cdu kip linkspartei

46: CDU, Spahn

Bild.de : cdu spahn maiziėr politik altmaier peter thoma jen innenminist tillich geißler generalsekretär ziemiak sachsen kretschmer ministerpräsid oppermann tauber minist partei

focus.de : cdu spahn tauber politik jen peter generalsekretär altmaier kretschmer tillich bosbach sachsen ditfurth partei jahren schäubl wolfgang union maiziėr amt

spiegel.de : cdu spahn politik altmaier sachsen partei jen generalsekretär tillich peter kretschmer geißler ministerpräsid kritik jahren tauber paul günther union klöckner

stern.de : cdu spahn geißler generalsekretär politik jen peter tauber heiner twitter jahren altmaier partei gysi berlin tot rheinland alter union kretschmer

tagesschau.de : cdu spahn jen kabinett tillich kretschmer politik tauber sachsen berlin präsidiumsmitgli peter generalsekretär günther partei michael jahren ministerpräsid staatssekretär wahlkampf

welt.de : cdu spahn altmaier jen peter politik welt tauber generalsekretär kretschmer geißler partei laschet kanzleramt sachsen ministerpräsid sonntag maiziėr amt politisch

zeit.de : cdu spahn sachsen tillich politik geißler generalsekretär jen kretschmer altmaier partei heiner michael peter jahren ministerpräsid finanzministerium politisch stanislaw union

A.2 Regression Results

A.3 Cross Correlation Coefficient

	topic_name	parameter	Estimate	Std. Error	t value	p
1	1: SPD, M.Schulz	(Intercept)	0.037	0.003	11.504	0.000
2	1: SPD, M.Schulz	FOCUS ONLINE	-0.008	0.004	-2.045	0.041
3	1: SPD, M.Schulz	SPIEGEL ONLINE	-0.008	0.004	-1.878	0.060
4	1: SPD, M.Schulz	stern.de	-0.009	0.004	-2.482	0.013
5	1: SPD, M.Schulz	Tagesschau.de	-0.015	0.004	-3.423	0.001
6	1: SPD, M.Schulz	DIE WELT	-0.015	0.004	-3.829	0.000
7	1: SPD, M.Schulz	ZEIT ONLINE	-0.012	0.004	-2.760	0.006
8	2: B90/ Die Grünen	(Intercept)	0.018	0.003	6.420	0.000
9	2: B90/ Die Grünen	FOCUS ONLINE	-0.005	0.003	-1.462	0.144
10	2: B90/ Die Grünen	SPIEGEL ONLINE	0.001	0.004	0.295	0.768
11	2: B90/ Die Grünen	stern.de	0.006	0.003	1.606	0.108
12	2: B90/ Die Grünen	Tagesschau.de	-0.003	0.004	-0.716	0.474
13	2: B90/ Die Grünen	DIE WELT	-0.004	0.003	-1.121	0.262
14	2: B90/ Die Grünen	ZEIT ONLINE	0.004	0.004	1.026	0.305
15	4: Great Coalition debates	(Intercept)	0.064	0.005	13.476	0.000
16	4: Great Coalition debates	FOCUS ONLINE	-0.010	0.006	-1.846	0.065
17	4: Great Coalition debates	SPIEGEL ONLINE	0.002	0.006	0.327	0.744
18	4: Great Coalition debates	stern.de	-0.033	0.006	-5.758	0.000
19	4: Great Coalition debates	Tagesschau.de	-0.006	0.006	-0.927	0.354
20	4: Great Coalition debates	DIE WELT	-0.013	0.006	-2.386	0.017
21	4: Great Coalition debates	ZEIT ONLINE	0.001	0.007	0.197	0.843
22	10: M.Schulz, vs. A.Merkel	(Intercept)	0.009	0.002	3.848	0.000
23	10: M.Schulz, vs. A.Merkel	FOCUS ONLINE	0.000	0.003	0.178	0.859
24	10: M.Schulz, vs. A.Merkel	SPIEGEL ONLINE	0.002	0.003	0.537	0.591
25	10: M.Schulz, vs. A.Merkel	stern.de	0.007	0.003	2.565	0.010
26	10: M.Schulz, vs. A.Merkel	Tagesschau.de	0.003	0.003	0.874	0.382
27	10: M.Schulz, vs. A.Merkel	DIE WELT	-0.003	0.003	-1.095	0.273
28	10: M.Schulz, vs. A.Merkel	ZEIT ONLINE	0.011	0.003	3.599	0.000
29	13: A.Merkel, election campaign	(Intercept)	0.024	0.003	9.272	0.000
30	13: A.Merkel, election campaign	FOCUS ONLINE	0.002	0.003	0.470	0.638
31	13: A.Merkel, election campaign	SPIEGEL ONLINE	0.003	0.003	0.921	0.357
32	13: A.Merkel, election campaign	stern.de	0.008	0.003	2.724	0.006
33	13: A.Merkel, election campaign	Tagesschau.de	-0.006	0.004	-1.578	0.115
34	13: A.Merkel, election campaign	DIE WELT	0.001	0.003	0.328	0.743
35	13: A.Merkel, election campaign	ZEIT ONLINE	0.004	0.003	1.190	0.234
36	17: Debates within SPD	(Intercept)	0.035	0.003	10.158	0.000
37	17: Debates within SPD	FOCUS ONLINE	-0.006	0.004	-1.313	0.189
38	17: Debates within SPD	SPIEGEL ONLINE	-0.010	0.004	-2.172	0.030
39	17: Debates within SPD	stern.de	-0.002	0.004	-0.420	0.675
40	17: Debates within SPD	Tagesschau.de	-0.014	0.005	-2.995	0.003
41	17: Debates within SPD	DIE WELT	-0.011	0.004	-2.690	0.007
42	17: Debates within SPD	ZEIT ONLINE	-0.008	0.004	-1.738	0.082
43	20: AfD, right-wing radicalism	(Intercept)	0.018	0.003	6.252	0.000
44	20: AfD, right-wing radicalism	FOCUS ONLINE	-0.002	0.004	-0.684	0.494
45	20: AfD, right-wing radicalism	SPIEGEL ONLINE	-0.003	0.004	-0.800	0.424
46	20: AfD, right-wing radicalism	stern.de	-0.003	0.003	-0.899	0.369
47	20: AfD, right-wing radicalism	Tagesschau.de	-0.000	0.004	-0.029	0.977
48	20: AfD, right-wing radicalism	DIE WELT	0.002	0.003	0.489	0.625
49	20: AfD, right-wing radicalism	ZEIT ONLINE	-0.002	0.004	-0.567	0.571
50	22: SPD, stuffing debates	(Intercept)	0.018	0.003	7.054	0.000
51	22: SPD, stuffing debates	FOCUS ONLINE	-0.004	0.003	-1.083	0.279
52	22: SPD, stuffing debates	SPIEGEL ONLINE	0.010	0.003	3.089	0.002
53	22: SPD, stuffing debates	stern.de	-0.006	0.003	-1.984	0.047
54	22: SPD, stuffing debates	Tagesschau.de	-0.003	0.003	-0.913	0.361
55	22: SPD, stuffing debates	DIE WELT	-0.004	0.003	-1.225	0.221
56	22: SPD, stuffing debates	ZEIT ONLINE	0.005	0.003	1.435	0.151

	topic_name	parameter	Estimate	Std. Error	t value	p
57	23: CSU, Söder & Seehover	(Intercept)	0.035	0.004	9.361	0.000
58	23: CSU, Söder & Seehover	FOCUS ONLINE	-0.003	0.005	-0.739	0.460
59	23: CSU, Söder & Seehover	SPIEGEL ONLINE	0.006	0.005	1.286	0.199
60	23: CSU, Söder & Seehover	stern.de	-0.002	0.004	-0.358	0.720
61	23: CSU, Söder & Seehover	Tagesschau.de	-0.006	0.005	-1.132	0.257
62	23: CSU, Söder & Seehover	DIE WELT	-0.005	0.005	-1.157	0.247
63	23: CSU, Söder & Seehover	ZEIT ONLINE	0.002	0.005	0.437	0.662
64	26: Jamaica fail, Reelections or GroKo?	(Intercept)	0.035	0.003	12.152	0.000
65	26: Jamaica fail, Reelections or GroKo?	FOCUS ONLINE	-0.004	0.004	-1.174	0.240
66	26: Jamaica fail, Reelections or GroKo?	SPIEGEL ONLINE	-0.004	0.004	-1.121	0.262
67	26: Jamaica fail, Reelections or GroKo?	stern.de	-0.012	0.003	-3.560	0.000
68	26: Jamaica fail, Reelections or GroKo?	Tagesschau.de	-0.010	0.004	-2.608	0.009
69	26: Jamaica fail, Reelections or GroKo?	DIE WELT	-0.012	0.004	-3.374	0.001
70	26: Jamaica fail, Reelections or GroKo?	ZEIT ONLINE	-0.006	0.004	-1.442	0.149
71	27: Jamaica Coalition debates	(Intercept)	0.066	0.005	14.338	0.000
72	27: Jamaica Coalition debates	FOCUS ONLINE	-0.029	0.006	-5.204	0.000
73	27: Jamaica Coalition debates	SPIEGEL ONLINE	-0.009	0.006	-1.400	0.161
74	27: Jamaica Coalition debates	stern.de	-0.028	0.005	-5.130	0.000
75	27: Jamaica Coalition debates	Tagesschau.de	-0.018	0.006	-2.752	0.006
76	27: Jamaica Coalition debates	DIE WELT	-0.007	0.005	-1.376	0.169
77	27: Jamaica Coalition debates	ZEIT ONLINE	-0.003	0.006	-0.509	0.611
78	30: AfD, F.Petry & Meuthen	(Intercept)	0.026	0.003	7.873	0.000
79	30: AfD, F.Petry & Meuthen	FOCUS ONLINE	-0.005	0.004	-1.248	0.212
80	30: AfD, F.Petry & Meuthen	SPIEGEL ONLINE	0.002	0.004	0.490	0.624
81	30: AfD, F.Petry & Meuthen	stern.de	-0.004	0.004	-0.950	0.342
82	30: AfD, F.Petry & Meuthen	Tagesschau.de	-0.014	0.004	-3.234	0.001
83	30: AfD, F.Petry & Meuthen	DIE WELT	-0.011	0.004	-2.640	0.008
84	30: AfD, F.Petry & Meuthen	ZEIT ONLINE	0.004	0.004	0.918	0.359
85	32: AfD, Gauland & Weidel	(Intercept)	0.023	0.003	6.837	0.000
86	32: AfD, Gauland & Weidel	FOCUS ONLINE	0.001	0.004	0.141	0.888
87	32: AfD, Gauland & Weidel	SPIEGEL ONLINE	-0.001	0.004	-0.170	0.865
88	32: AfD, Gauland & Weidel	stern.de	0.007	0.004	1.627	0.104
89	32: AfD, Gauland & Weidel	Tagesschau.de	-0.006	0.005	-1.371	0.170
90	32: AfD, Gauland & Weidel	DIE WELT	0.006	0.004	1.494	0.135
91	32: AfD, Gauland & Weidel	ZEIT ONLINE	0.006	0.005	1.246	0.213
92	37: AfD & DIE LINKE in parliament	(Intercept)	0.025	0.003	7.013	0.000
93	37: AfD & DIE LINKE in parliament	FOCUS ONLINE	0.005	0.004	1.159	0.246
94	37: AfD & DIE LINKE in parliament	SPIEGEL ONLINE	0.013	0.005	2.827	0.005
95	37: AfD & DIE LINKE in parliament	stern.de	0.001	0.004	0.190	0.850
96	37: AfD & DIE LINKE in parliament	Tagesschau.de	0.005	0.005	1.089	0.276
97	37: AfD & DIE LINKE in parliament	DIE WELT	0.002	0.004	0.369	0.712
98	37: AfD & DIE LINKE in parliament	ZEIT ONLINE	0.010	0.005	2.128	0.033
99	46: CDU	(Intercept)	0.019	0.003	7.348	0.000
100	46: CDU	FOCUS ONLINE	-0.001	0.003	-0.321	0.748
101	46: CDU	SPIEGEL ONLINE	0.002	0.003	0.707	0.479
102	46: CDU	stern.de	-0.006	0.003	-1.979	0.048
103	46: CDU	Tagesschau.de	-0.011	0.003	-3.373	0.001
104	46: CDU	DIE WELT	-0.002	0.003	-0.523	0.601
105	46: CDU	ZEIT ONLINE	-0.000	0.004	-0.085	0.932

Table 2: Cross Correlation at lag 0

	Var1	AfD	FDP	Grüne	Linke	SPD	Union
1	1: SPD, M.Schulz	-0.636	0.072	-0.438	-0.693	0.752	0.199
2	10: M.Schulz, vs. A.Merkel	0.438	0.566	0.505	0.341	-0.229	-0.758
3	13: A.Merkel, election campaign	0.098	0.350	0.440	-0.060	0.076	-0.507
4	17: Debates within SPD	-0.770	0.075	-0.676	-0.754	0.841	0.382
5	2: B90/ Die Grünen	0.253	-0.092	0.185	0.480	-0.329	-0.048
6	20: AfD, right-wing radicalism	0.225	0.369	0.224	0.037	-0.157	-0.319
7	22: SPD, stuffing debates	-0.710	-0.112	-0.465	-0.797	0.877	0.302
8	23: CSU, Söder & Seehover	-0.169	-0.897	-0.249	-0.283	0.145	0.565
9	26: Jamaica fail, Reelections or GroKo?	-0.251	-0.501	-0.388	-0.258	0.047	0.537
10	27: Jamaica Coalition debates	-0.197	-0.439	-0.266	-0.222	0.000	0.436
11	30: AfD, F.Petry & Meuthen	0.410	-0.032	0.351	0.296	-0.348	-0.224
12	32: AfD, Gauland & Weidel	-0.220	0.430	0.221	-0.315	0.336	-0.296
13	37: AfD & DIE LINKE in parliament	-0.217	-0.521	-0.418	-0.003	0.238	0.552
14	4: Great Coalition debates	-0.165	0.336	-0.361	0.292	-0.126	0.228
15	46: CDU	-0.083	-0.066	-0.108	-0.286	0.285	-0.007