

Analysing News about political Parties in German Online Media

Konstantin-Julius Oellers¹, Christian Wirtgen¹

¹*Master Informatik, Hochschule RheinMain, Wiesbaden, Germany*

konstantin-julius.oellers@student.hs-rm.de, christian.wirtgen@student.hs-rm.de

Keywords: Data Analytic; News; German Parties; Online Media

Abstract: The political environment in Germany has undergone major changes in recent years. The major popular parties, the liberal-conservative Union (CDU/CSU) and the social-democratic SPD, are losing their votes to the smaller and newer parties like the Greens (Alliance 90/The Greens) and the right-wing Alternative for Germany (AfD). The influence of the new media is often cited as one of the main reasons for this development. The new media are often accused of being more susceptible to manipulation and false reports than the classic media. The role of these media has been the subject of much debate for some time now: Through selective reporting and public discussions, the media should have an influence on political events and the formation of public opinion. This paper examines this allegation by analyzing how various parties are being reported. For this purpose, all articles about political parties were collected by selected media over a period of 5 years. This data were used for different analyses like Sentiment Detection, Language Analysis and Data Clustering. For example it was found that AfD articles have on average the longest length and have the highest negative polarity score of all parties. During the clustering it could be observed that all parties are reported on most frequently at the provincial and Federal state level. Based on the Pearson correlation coefficient it could be shown that there is no correlation between the article scoring and survey values or Google Trend for the SPD und Green Party. In contrast, the AfD party has a high correlation between the amount of published articles and survey values. When analysing certain events, it can be observed that attacks, whether Islamist or right-wing, have no influence on the trend of the AfD party. No causality could be established between coverage and survey results.

1 INTRODUCTION

With the rise of the Internet, the way of reporting has changed. While the number of daily newspaper subscriptions has been declining for several years, the number of people who obtain their news via the Internet is increasing. This trend towards digitalisation was recognised by publishers at an early stage. The world's first news portal, Spiegel Online, was launched in Germany in 1994.^[spi] In addition to Spiegel, many other major daily newspapers and magazines now have offshoots in the form of online news portals. These websites still dominate the ranking of the most visited news sites today. Some news portals offer the possibility for readers to post comments on the articles on the site.

In the 2010s a political change has taken place in Germany. Since the founding of AfD in 2013, a strong downward trend of the major German parties CDU and SPD has become apparent. Apart from the AfD, smaller parties such as the Greens have benefited most from this plunge in voter favour. The rea-

sons for this political change are often seen in the high level of reporting on certain topics. The rise of the right-wing party AfD in particular is said to be connected with the high level of reporting. "There is no such thing as bad publicity" is a famous and controversial statement which highlights the influence of the media. It states that all mentions in media about a person or a topic make it more relevant and popular, even if they put it in a bad light. Because of the media's influence on the formation of public opinion, the press is often called the fourth power in the state.^[Fou]

In this paper, we try to examine this thesis on the basis of the coverage of political parties in the German online media. For the period from 1 January 2013 to 31 December 2019, all published news from certain online media about the German political parties AfD, Greens and SPD will be collected. All comments on AfD articles during this period will also be collected. In order to be able to distinguish benevolent from negative articles and comments, the data will be extended by a score. This score is determined with a text analysis. The collected and extended data

Website	Publisher	Confidence Rank	Visitor Rank	Comments	Greens Articles	SPD Articles	AfD Articles	Σ
Bild	Springer	18/18	1	-	28.427	40.534	7.196	76.157
Welt	Springer	-	6	✓	2.257	5.530	4.086	11.873
Spiegel	Spiegel	10/18	3	✓	17.341	19.171	3.617	40.129
Süddeutsche	SZ	5/18	14	-	90.234	103.809	21.698	215.741
Tagesschau	ARD	1/18	-	✓	1.047	2.880	988	4.915
t-online	Ströer	17/18	2	✓	36.110	42.201	12.632	89.631
Σ					174.104	214.125	50.226	438.455

Figure 1: Information table about the collected data of the websites

is then analysed using various methods. Among other things, the amount and score of the articles and comments is compared with the survey values and Google trends. These results are displayed graphically in diagrams and are also examined mathematically with a correlation coefficient. In order to cover the related topics in the media, the articles are assigned to a specific cluster. These different clusters are then visualized in statistics. Finally, the collected information and results are summarized and discussed.

2 DATA AND METHODS

We will now present the investigated websites and examine the data structure of the collected information. Then we will explain the methods used to extend and examine the data.

2.1 Data and Terminology

In order to obtain a heterogeneous picture of the German media landscape, very different websites were examined. As can be seen on figure 1, all of the websites examined, with the exception of the public-law Tagesschau.de, are owned by private publishers. A total of 438.455 news articles and 655.079 comments were collected. All data sets together comprise hard disk space of 1.6 GB. The confidence ranking is based on a survey from 2018.^[con] The visitor rank is based on a data evaluation from January 2020.^[ran]

Some of the news portals offer the possibility to search all articles with a specific day. However, in order to be able to adequately compare the data between the different websites, a uniform criterion must be defined when an article reports on a political party. The only practicable solution is full-text search, which is offered by each news site. Any article that contains the word "SPD", "Greens" or "AfD" in the text will be considered as an article about the party.

Name	Datatype	Optional
Author	String	-
Headline	String	-
Headline intro	String	✓
Datetime	Date	-
hasComments	Boolean	-
URL	String	-
Article intro	String	-
Tags	Array of Strings	✓
Comments	Array of Objects	✓
Category	Array of Strings	✓
Search query	String	-
Source	String	-
Score	Float	-

Figure 2: Data structure of the news articles

Accordingly, articles can occur several times. It is also possible that an article does not mention a political party at all, but refers to a completely different word. Especially the word "green" can be used in the wrong context. For some media it was possible to remove non-political articles using the URL. This reduces the noise significantly.

Name	Datatype	Optional
Author	String	-
Datetime	Date	-
Likes	Boolean	✓
Score	Float	-

Figure 3: Data structure of the comment object

For some of the articles in Spiegel, Welt, Tagesschau and t-online it is possible for users to write comments directly on the website. Since there is an enormous number of comments, only comments on articles about AfD were examined. However, the comment function was deactivated for all AfD articles on Tagesschau.de. Since the effort to collect

the comments at t-online proved to be enormously time-consuming, it was also excluded. Therefore only the comments of Spiegel and Welt can be examined. However, for each article on every website the information is stored whether it contains comments. Figure 2 shows the data structure of the collected items. The data structure of the comments is explained in Figure 3.

In order to trace the parties' popularity chronologically, two additional data sources were used. One is the archive of survey values (so-called Sunday polls), which are conducted at regular intervals by various renowned polling institutes. The other data source is Google Trends. There, the number of search queries to the Google search engine can be tracked precisely. The collected data is stored in a MongoDB.

2.2 Sentiment Detection

A sentiment analysis was applied to all collected articles. For this purpose, the text of the article was converted into word tokens. In order to form word marks, all special characters have been removed from the article text. After this the words must first be reduced to their root form and so the lemma of the word can be identified (Stemming and Lemmatization step). This was done with the NLP library *spaCy*^[HM17]^[lib]¹.

A sentiment score was then calculated using the *SentiWS*^[RQH10]² list. In *SentiWS* (v2.0)² each word has a positive and negative polarity within the interval of [-1; 1]. By the summation of the polarity of all article words the score is calculated and it can be estimated whether an article is positive or negative.

2.3 Language Analysis

The language analysis examines the most commonly used words about the parties. The basis for this is the content of the articles. First, all punctuation marks must be removed from the text. Then certain German stop words were removed. The 20 most frequently used words are then combined in a word cloud. The larger a word is in the word cloud, the more often it appears in the text.

2.4 Examining Survey Values and Google Trends

In order to compare the number of news articles with the poll results and Google Trends of a party, the arti-

cles about a examined party between two dates were examined in different four modes:

- The percentage of reporting about the party was examined. For this, the number of articles about the party was divided by the number of articles collected in the period.
- The score was examined by calculating the average score during this period.
- The average number of comments was analysed.
- The average score of comments was analysed.

In order to visualize the data, we mapped the results into a statistic with two differently scaled Y-axes: The survey values / Google Trends are drawn as a line chart and the examined amount of articles and average score as a stacked bar chart. The X-axis represents the time period under investigation. The two federal elections of 2013 and 2017 are marked in the statistics. Since only AfD comments are collected, the last two modes are limited to these comments only.

To show a linear correlation between the survey values / Google Trends and the articles we have used the Pearson correlation coefficient. The result is a value between -1 and +1, where +1 means total positive linear correlation, 0 means no linear correlation and -1 is a total negative linear correlation.

2.5 Data Clustering

Text clustering is used to find out which topics predominate in articles, how often these topics occur and how they relate to each other. As feature extraction *TF-IDF*^[R+03] is used. TF-IDF stands for Frequency-Inverse Document Frequency and uses a numerical statistic to show how important a word is to its text corpus. Uninteresting words like pronouns or indefinite articles are first filtered out using a stop-word list. The actual clustering happens in the next step with the unsupervised learning algorithm *k-means++*^[AV06]. K-Means clustering tries to find cluster centers that are representative of certain regions of the data. K-means++ is a extended version of kmeans with a randomized seeding technique which improves the speed and the accuracy. Both k-means algorithm require the number of clusters as a free parameter.

To identify a reasonable number of clusters the Error Sum of Squares (SSE)^[Ste07] was calculated for a test data set. This test data set consists of 5000 randomly selected articles from the entire data set. The SSE describes the squared differences between each observation and its group's mean. Therefore a low SSE value is searched for with a high cluster number. Figure 4 illustrates the SSE graph for the test data set.

¹<https://github.com/explosion/spaCy>

²<https://wortschatz.informatik.uni-leipzig.de/en>

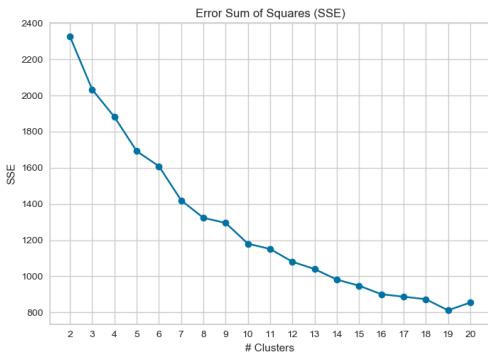


Figure 4: Error Sums of Squares for test data set

As cluster number 16 was chosen (with a SSE value of 940). A larger number of clusters would cause the SSE to decrease further, but then there would be difficulties with the visualization.

The feature extraction and the clustering was implemented with the *sklearn* library^[BLB⁺13]³. After clustering, the distribution of the clusters and the top terms per cluster is determined. As final step the results are presented as a *t-SNE Corpus Visualization*^[VDM14] from the Yellowbrick-API⁴. It was also attempted to use the DBSCAN^[BK07] algorithm to cluster the data sets, but the results were poorer or unusable and running time was enormously long, so the K-means continued to be used. There were three data sets in total. One data set contains all articles in all media about a particular party.

3 RESULTS

Based on the methods presented, the collected data were examined. The results of the investigations are described in this section.

3.1 Articles Lengths

Figure 5 shows the average length of the articles content, grouped by medium and party. The length is the total number of characters in a article used. No characters were filtered out. The average value was calculated by adding up the article length and dividing it by the number of according articles. The global average length for all articles is 2487. It can be seen that Spiegel and Welt publish by far the longest articles.

³<https://github.com/scikit-learn/scikit-learn>

⁴<https://github.com/DistrictDataLabs/yellowbrick/>

The Bild, on the other hand, clearly writes the shortest articles and is below the global average. The other media are near the average. Looking at the length of articles in relation to the parties, it is noticeable that the AfD receives surprisingly long articles compared to the other two parties. This can be seen, for example, at t-online and Bild, where an AfD article contains almost 200-400 characters more. The AfD also has the longest articles length on the global average with a value of 2652.

Website	Green	SPD	AfD
Bild	1629	1482	1878
Welt	4201	4437	3927
Spiegel	4019	3550	4086
Süddeutsche	2676	2391	2435
Tagesschau	2968	2912	3117
t-online	2359	2268	2607
Global	2597	2358	2652

Figure 5: Mean articles lengths

The SPD on contrast has shorter articles in comparison, this is particularly noticeable at Spiegel, where the SPD has on average over 500 characters less. With a global mean value of 2597, the Greens rank between the AfD and the SPD. The article length is very balanced for Süddeutsche and Tagesschau.

3.2 Clustered Data

When analyzing the different data sets, it was noticed that clusters are always formed on certain topics. A total of 6 different cluster types have been identified. For example, there is the cluster type C1, which is clearly related to foreign policy or cluster type C3 which contains articles at a provincial level. Each cluster type has typical words, that describe its type of cluster. Figure 6 shows this cluster types with their most common words. Almost every cluster type occurs in every data set.

The first clustering set contains all articles of the AfD party. Figure 24 shows the cluster distribution sorted according to the cluster size and the top ten words of each cluster. It is striking that in the top words the party AfD is mentioned in connection with "Polizei", "Pegida" and "Verfassungsschutz" and "Deutsch". It can be seen that various clusters have formed at the provincial level cluster type. Here are often mentioned different federal states like "Sachsen", "Thüringen" or "Bayern". In the C2

Cluster Type	Cluster Topic	Typical Words
C1	Foreign policy	"Außenminister", "USA", "Russland", "Botschaft", "US"
C2	Domestic policy	"Bund", "Innenminister", "Berlin", "Deutschland", "Koalition"
C3	Provincial level	"Landtag", "Gemeinde", "Bürgermeister", "Stadtrat"
C4	European level	"EU", "Euro", "Brüssel", "Komission"
C5	Elections, polls and surveys	"Umfrage", "Stimmen", "Prozent", "Wahlkreis"
C6	Party near topics	"Partei", "Parteitag", "Mitglieder"

Figure 6: Cluster Types

clusters the AfD party is often mentioned in connection with other parties such as SPD and the Greens or concrete politicians like "Merkel". C5 is related to the Bundestag elections in 2017 or to other surveys. Cluster type C6 is very much concerned with specific AfD politicians or their members, such as "Petry", "Höcke", "Gauland" or "Meuthen".

The TSNE projection presented in Figure 27 shows that the clusters at provincial level are arranged at the top, above the left side to the bottom half. C4 Clusters are almost exclusively at the margin. The C6 party clusters are located on the right-hand edge and are directly adjacent to each other. C2 Clusters are distributed in the middle of the projection. The clusters regarding to survey or election (C6) themes are isolated from the remaining clusters.

Figure 7 displays the distribution of the cluster types for the AfD party. For each cluster type all related articles were summed up and divided by the total number to get the percentage value. It can be seen that AfD is clearly most frequently reported at the national level. Foreign policy is not reported on AfD at all. The remaining cluster types are evenly distributed in fact.

Types of clusters	Cluster Number	Percentage
C1	-	0.0%
C2	2, 12	14.7%
C3	0, 5, 9, 11, 13, 14, 15	44.3%
C4	1, 7	10.2%
C5	4, 8	17.2%
C6	6, 10	13.4%

Figure 7: Distribution of the cluster types in the AfD data set

Figure 25 shows the distribution of all SPD articles. It is noticeable that the party SPD is often associated with "Koalition", "Merkel", "Berlin" and "CDU". In the C3 clusters city names like "Berlin" or "München" are often mentioned. In C1 clusters other states are named ("Russland", "USA", "Türkei") and

the Federal Foreign Ministers ("Gabriel", "Steinmeier"). C6 is mainly about the coalition with the CDU and names well-known politicians from both sides ("Merkel", "Schulz", "Nahles"). In the TSBE projection (figure 26) it can be seen that the cluster of foreign policy is located at the lower half of the edge. By contrast, the clusters of domestic policy are on the left or in the middle of the projection. The C3 clusters are distributed over the entire projection. Similar to Figure 27, the clusters about surveys and elections are located on the edge or outside. The C4 Clusters are at the margin bottom. The distribution in Figure 3 shows that most of the articles can be assigned to the country level. However, at just under 28%, there is also a very large domestic part at national level. Articles on foreign policy, Europe or party internals play only a subordinate role in reporting on the SPD.

Types of clusters	Cluster Number	Percentage
C1	6	7.6%
C2	5, 7, 13, 15	28.4%
C3	0, 1, 3, 4, 8, 11	42.8%
C4	12	4.4%
C5	2, 9, 10	12.5%
C6	14	4.1%

Figure 8: Distribution of the cluster types in the SPD data set

The last data set contains all Green articles and is shown in figure 28. The green party is mentioned in connection with "EU", "CO2", "klimaneutral" and with all major German parties. It is noticeable that in many different clusters often words like "Euro", "Kosten", "Millionen" or "Milliarden" are mentioned. The largest clusters can be assigned to the provincial level (C3). It is surprising that Cluster 9 is a very small cluster with a very concrete topic about "Maut", "PKW" and "EUGH" and belongs to C2. With cluster 6 there is another concrete cluster, which is related to "bagdad", "raketen", "tesla" and "us" and can be assigned to foreign policy (C1). As with the other data sets, there are clusters on survey and elections polls. As can be seen in the TSNE fig-

ure 29, the C3 clusters are arranged centrally from top to bottom. The clusters with European themes are distributed around them on both sides. The clusters regarding to survey or election themes are outside from the remaining clusters. It is interesting to note that the clusters on domestic policy are at the margins of the projection. The cluster on foreign policy is located outside and isolated in the upper area.

Figure 9 shows that C3 is by far the largest category for the Green party with 46.6%. This is followed by the domestic and European clusters with 16.3% and 26.6% respectively. Surprisingly, in polls and surveys reports the Green Party does not appear as often as the other two parties (Green: 9.4%, AfD: 17.2%, SPD: 12.5%). The Greens are very rare in foreign policy articles with only 0.9%.

Types of clusters	Cluster Number	Percentage
C1	6	0.9%
C2	8, 9, 15	16.3%
C3	0, 1, 3, 10, 12, 14	46.6%
C4	4, 7, 11, 13	26.6%
C5	2, 5	9.4%
C6	-	0.0%

Figure 9: Distribution of the cluster types in the Green data set

In summary, it can be observed that all parties are reported on most frequently at the provincial and Federal state level. Each party has typical keywords with which they are associated. Surprisingly, the word refugee does not appear in any of the AfD clusters, although the word is strongly associated with the party. The TSNE projections show that the cluster types C2 and C3 are situated close together and have a high similarity. The cluster with foreign reference (C1, C4) are more on the edge of projection and are therefore less similar to other clusters.

3.3 Scoring

As described in chapter 2.2 a score was calculated for each article, which indicates whether an article is positive or negative. The data set shows that on average all parties are reported negatively. The best value is achieved by the Green Party with an overall average value of -0.687. The SPD follows with a value of -0.793. The AfD party is reported to be the most negative with a total mean value of -1.296. Figure 10 shows the average score of a party according to the media sources. AfD comes off badly in all media, with values between -0.94 (Bild) and -2.239 (Spiegel). It can be noted that AfD is reported almost

twice as negatively as other parties. A closer look at the media sources reveals that the Bild is the most balanced in their coverage of the parties, followed by Süddeutsche. All other media reported more negatively on the parties with a factor around of 2 to 3. It stands out that Spiegel Online has the most negative reporting compared to other media, followed by Tagesschau and Welt.

Website	Green	SPD	AfD
Bild	-0.496	-0.563	-0.941
Welt	-1.170	-1.105	-1.960
Spiegel	-1.340	-1.643	-2.249
Süddeutsche	-0.483	-0.673	-1.063
Tagesschau	-1.238	-1.297	-1.598
t-online	-0.998	-0.846	-1.386

Figure 10: Mean Score of a party subdivided according to media

Figure 38 shows the survey results in chronological comparison with the scoring of the AfD articles. From its inception in February 2013 until the end of 2014 the score is still relatively high. Only after that the score did fall and at the same time the survey results. Since the summer of 2015, the survey values have risen rapidly, while the score has continued to fall. Apart from a few extreme outliers, such as in 2016 and 2017, the score remains consistently very low. The survey values, on the other hand, have not fallen, but in some cases have risen. A comparison of the score with Google Trends in Figure 41 leads to a similar result.

A comparison with the SPD articles in Figure 37 shows that a falling score also leads to an increase in the survey values. However, this effect is significantly weaker than with AfD. Thus, the score and the survey values are quite constant over the entire period. Surprisingly, when comparing with Google Trends in Figure 40, there is a slight correlation between increased trends and increased survey values.

By contrast, the results of the analysis of the Green articles are quite different: when looking at Figure 36, almost no correlation between the survey values and the score can be found. The same applies to Google Trends in Figure 39.

Next, the correlation between the score and survey values and Google Trends was analyzed. Here, a positive correlation coefficient means a correlation between rising survey scores and rising score. A negative correlation coefficient means rising survey scores with falling score. As can be seen in Figure 11, a high negative correlation coefficient was found for the AfD. A low score for the AfD correlates for all me-

dia except for Süddeutsche both with increased survey values and higher Google trends. In the case of the Greens, Figure 12 shows a mixed result: some of the media show a low positive correlation, some other media a low negative correlation. Figure 13 shows a similarly mixed relationship for the SPD as for the Greens. However, the correlation coefficients of the Süddeutsche and t-online to the SPD's survey values are significantly lower than for the other media.

Website	Sonntagsfrage	Google Trends
Bild	-10.06%	-11.68%
Welt	-33.64%	-3.29%
Spiegel	-21.14%	-32.19%
Süddeutsche	4.48%	-3.13%
Tagesschau	-34.26%	-15.53%
t-online	-41.83%	-11.39%
Ø	-22.74%	-12.87%

Figure 11: Pearson correlation coefficient for the article scoring of the AfD

Website	Sonntagsfrage	Google Trends
Bild	4%	3.29%
Welt	8.27%	4.68%
Spiegel	-9.91%	12.26%
Süddeutsche	2.01%	-6.44%
Tagesschau	-16.8%	8.72%
t-online	7.47%	-0.82%
Ø	-0.82%	3.62%

Figure 12: Pearson correlation coefficient for the article scoring of the Greens

Website	Survey value	Google Trends
Bild	4.41%	2.17%
Welt	2.43%	4.75%
Spiegel	3%	-5.75%
Süddeutsche	-38.47%	13.96%
Tagesschau	4.63%	6.32%
t-online	-23.99%	11.3%
Ø	-8%	5.46%

Figure 13: Pearson correlation coefficient for the article scoring of the SPD

In the appendix are the visualizations for the scoring of the parties: Figure 38 for the AfD, Figure 37 for the SPD and Figure 36 for the Greens shows the statistics for the survey scores. Figure 41 for the AfD, Figure 40 for the SPD and Figure 39 for the Greens shows the statistics for Google Trends.

3.4 Language Analysis

A word cloud was created for each party with a generator⁵. Unimportant words were filtered out as described in chapter 2.5. As expected, the most frequently mentioned word in all parties was the party name itself. The other parties of the Bundestag also dominate the top list of the most used words.



Figure 14: Word cloud SPD

The Figure 14 shows the most used words in articles about the SPD. "Deutschland" (Germany) was mentioned most frequently in SPD articles on the 6th rank. The word "EU" ranks 10th. Angela Merkel and "AfD" only follow in 12th and 13th place.



Figure 15: Word cloud Greens

In the Green word cloud, which can be seen in Figure 15, the words "SPD" and "Prozent" (percent) were used most often. Although the word "EU" is only in 12th rank among the Greens, the word "Euro" is in 5th rank. The AfD is in 13th place and Merkel is not in the top 15 list.

⁵https://github.com/amueller/word_cloud



Figure 16: Word cloud AfD

In the AfD word cloud from figure 16, the second and third rank are the same as for the Greens. Surprisingly, "Deutschland" is in 8th rank, still behind "Grünen" in 7th rank. "Merkel" is on the the 10th rank in the afd word cloud.

3.5 Article Count

The statistics show very different results for the various parties. When looking at the results of AfD at figure 32, it is noticeable that initially, despite increased survey results, little was reported about AfD. It was not until the end of 2015 that reporting and survey results rose enormously. This high level of reporting reached its peak in early 2016. In 2017, both the number of articles and the survey results fell. In 2018 and 2019, however, both values rose again and remained at a constant level. Figure 32 shows a very similar results: Every month that Google's trends on AfD have increased, the number of articles has also increased.

Website	Sonntagsfrage	Google Trends
Bild	71.57%	41.1%
Welt	44.71%	21.47%
Spiegel	66.1%	32.59%
Süddeutsche	69.51%	29.95%
Tagesschau	40.07%	30.19%
t-online	70.54%	38.93%
Ø	60.42%	32.37%

Figure 17: Pearson correlation coefficient for the article count of the AfD

For the Greens, on the other hand, the proportion of reporting and survey values has remained fairly constant over the period 2013 to 2018. This can be seen in figure 30. From the end of 2018, however, the survey results increased enormously, although the

share in reporting has only increased slightly. The figure 33 shows that Google Trends and average amount of articles for the Greens remained also very constant.

Website	Sonntagsfrage	Google Trends
Bild	37.05%	8.69%
Welt	1.31%	22.51%
Spiegel	19.25%	16.83%
Süddeutsche	20.11%	-10.14%
Tagesschau	2.65%	-1.72%
t-online	25.86%	2.46%
Ø	17.7%	6.43%

Figure 18: Pearson correlation coefficient for the article count of the SPD

Figure 31 shows that the average proportion of reporting for the SPD has remained fairly constant over the entire period. The *Tagesschau* reported particularly frequently on the SPD. Here, too, there is a correlation between the increasing share of reporting on the SPD and rising poll ratings. Figure 34 shows that this also applies to Google Trends. However, this effect is weaker for both the survey values and Google Trends than for the AfD.

Website	Sonntagsfrage	Google Trends
Bild	3.07%	8.93%
Welt	-6.91%	-12.57%
Spiegel	-4.27%	7.37%
Süddeutsche	0.94%	4.71%
Tagesschau	-2.67%	12.72%
t-online	12.62%	1.27%
Ø	0.47%	3.74%

Figure 19: Pearson correlation coefficient for the article count of the Greens

This correlation between increasing number of comments and increasing survey values or Google Trends could be demonstrated by the correlation coefficients in Figure 17. With a value of 60.42% the average correlation coefficient of the survey results is almost twice as high as that of the Goolge Trends. The correlation coefficient, which is shown in figure 19, is for both survey values and Google Trends nearly zero. This means that there is almost no measurable correlation between the amount of articles about the Greens and their survey values or Google Trends.

3.6 Analysing specific Events

One question in this paper is if a specific event can have an influence on the trend of a party. To answer

this question, three different events were examined:

- Attack on the Berlin Christmas Market (2016)
- Attack on Synagogue in Halle (2019)
- Nomination of Martin Schulz (2017)

The first event is an Islamist attack in Berlin in 2016. The terrorist Anis Amri killed 12 and injured 55 people on December 19, by crashing a truck into a Christmas market^[ber]. The incident triggered a major media reaction and a discussion about public safety in Germany developed. The second attack, on the other hand, was committed by a right-wing extremist. According to investigators, the German Stephan Balliet intended to enter the synagogue in Halle by force of arms in order to kill people gathered there. The attempt failed, then the perpetrator attacked random people and killed 2 people and injured 3 others^[hal]. This incident caused huge dismay in politics, in the media and in various associations. The attack started a debate about antisemitism and right-wing networks in Germany.

Figure 43 shows the period before and after the attack from 05.12.16 to 15.01.17. The actual attack at the Christmas Market happened on 19. December. Although this event has been widely and often reported, the number of all articles published on this day is very small and climbs again in the following days. However, as can be seen from the blue graph, that the trend towards the AfD party is about 30% in the days before the attack and increases significantly on the day of the attack and the following day. In the next few days, the trend rapidly declines again. Here the impact of the attack can be clearly seen. Figure 42 shows the same time period but with the summed scores of all AfD articles. While the score was below -5 in the days before the attack, it drops significantly to above -15 after the attack. If the correlation coefficient is taken into account (figure 20, the overall influence of the attack is rather low.

Website	Article Count	Score
Bild	-1.24%	-1.88%
Welt	-14.12%	2.93%
Spiegel	12.8%	-14.89%
Süddeutsche	-0.74%	-7.25%
Tagesschau	-4.38%	4.61%
t-online	-5.78%	4.61%
Ø	-2.24%	-4.64%

Figure 20: Pearson correlation coefficient for the article count/Trend and article score/Trend of AfD in the period from 01.12.16 to 15.01.17

The second attack is illustrated by figure 44. In September and early October the trend value for the AfD party fluctuates between 30 and 40 percent. On the day of the attack (9. October) the trend rises to about 45% and then quickly falls back to just over 30%. A high peak can be observed from 25 October onwards what can be explained by the state elections in Thuringia^[Lan]. The number of published articles on AfD is low on the day of the attack and increases slightly in the following days. Figure 45 shows the average score for the articles in this time period. In the time before the attack the score is negative with slight deviations. On the day of the attack and the following days the score is constantly negative.

If one considers figure 21 the correlation coefficient for the article count and Google Trend is 18.74%. Thereby the correlation coefficient for the article score and Google Trend is only low with 3.14%. The difference between these two values can be explained by the state election in Thuringia, which influenced both the trend and the number of articles published. The actual attack has no significant impact.

Website	Article Count	Score
Bild	13.63%	-6.66%
Welt	1.13%	-6.47%
Spiegel	15.71%	-2.97%
Süddeutsche	37.25%	14.58%
Tagesschau	9.98%	8.05%
t-online	34.74%	12.34%
Ø	18.74%	3.14%

Figure 21: Pearson correlation coefficient for the article count/Trend and article score/Trend of AfD in the period from 15.09.19 to 30.10.17

The last point is not a classical event, but describes the great upswing of German politician Martin Schulz (SPD). On 29 January 2017 Martin Schulz was unanimously nominated by the SPD party executives as candidate for chancellor for the federal elections^[Schal]. After this nomination the so-called "Schulz-Hype" or "Schulz-Zug" arose^[Schb]. In the weeks that followed, the SPD made significant gains in all polls nationwide and reported more than 10,000 new party entries^[SPD].

Figure 46 shows the time around the nomination of Martin Schulz with the survey results and percentage of reporting for the SPD party. In the weeks prior to the nomination, the SPD had a poll result of slightly over 20%. In the week before and after the nomination the survey results rise to over 30% and remain constantly high in the following weeks. The number of published articles increases during this time. Fig-

ure 47 shows the same time period but with the average score value. It can be seen that the score in the last two weeks of February increases significantly and even reaches a positive value. In the weeks that followed, the reporting became more negative again.

Google Trend compared to the article count for the SPD party is shown in figure 48 in the period from 01.01.17 to 30.03.17. The trend for the SPD fluctuates between 40 and 50 percent in January 2017. On January 24, the value spikes high upwards. This can be explained by the announcement that Sigmar Gabriel renounced his candidacy for chancellor and spoke out in favour of Schulz as the new top candidate and SPD leader^[Gab]. The next big spike is on January 19, when Martin Schulz was unanimously elected as candidate for chancellor by the SPD party executive. After the nomination, the trend value fluctuates between 50 and 60 percent in the following weeks. In March, the trend continues to fall steadily towards 40-50%. Furthermore one can see that the number of published articles about the SPD increases after the nomination. Figure 49 shows the score average score compared to Google Trends in the same time period. It can be seen that the score in early February is always negative and increases significantly and even reaches a positive value. In the weeks that followed, the reporting became more negative again. After the announcement that Sigmar Gabriel will not be a candidate for chancellor, the reporting about the SPD becomes much more positive and even reaches positive values.

Website	Sonntagsfrage	Google Trends
Bild	59.5%	20.56%
Welt	41.42%	35.65%
Spiegel	40.547%	3.06%
Süddeutsche	79.66%	22.13%
Tagesschau	15.19%	5.57%
t-online	21.6%	12.64%
Ø	42.99%	16.6%

Figure 22: Pearson correlation coefficient for the article count of SPD in the from 01.01.17 to 30.03.17

The Pearson correlation coefficient for the article scoring is shown in figure 22. It can be clearly seen that there is a very strong correlation between the increasing survey values and the increasing number of published articles with value of 42.99%. Each single medium has a high to very high positive coefficient. The correlation between Google Trend and the number of articles is lower with 16.6%. With a value of 36.89 percent, the correlation between survey value

Website	Sonntagsfrage	Google Trends
Bild	50.49%	7.88%
Welt	69.24%	11.54%
Spiegel	34.7%	-13.54%
Süddeutsche	41.27%	19.5%
Tagesschau	3.78%	5.57%
t-online	21.88%	16.29%
Ø	36.89%	9.18%

Figure 23: Pearson correlation coefficient for the article scoring of the SPD in the time from 01.01.17 to 30.03.17

and article rating is also very high (figure 23). As well here, the coefficient for Google Trend is lower with 9.18%.

In summary, it can be observed that attacks, whether Islamist or right-wing, have no influence on the trend of the AfD party. Although the Google trend is increasing in the attacks for AfD, it has no correlation with the media examined. This probably means that people who are sympathetic to AfD switch to other media or social networks. The situation is different with the nomination of Martin Schulz as candidate for chancellor. Here the media examined show a high correlation with the Google trend and the number of articles published.

3.7 Analysing Comments

The commentary area of Spiegel and Welt differs mainly in its scope, as can be seen in figure 50 for the survey values and in figure 51 for Google Trends. In the case of Spiegel, the number of comments remained fairly constant over the entire period. At Welt.de, which only started its comment function at the end of 2016, the number of comments increased rapidly and quickly exceeded the number of Spiegel.de comments. The average of almost 600 comments per AfD article in February 2019 was a strong outlier. Since then, the number of comments at Welt.de has also stabilised.

At Welt.de a correlation of rounded 54.75% could be proven for the survey results and amount of comments. At Spiegel.de this value is only about 11.92%. A higher correlation was also shown for Welt.de compared with Google Trends. This correlation coefficient is around 17.63% at Welt.de and around 4.43% at Spiegel.de.

A comparison of the score of the comments with the survey values in Figure 52 and Google Trends in Figure 53 shows a similar result. With few exceptions the score is negative throughout. Here, a correlation between falling score and rising survey values and Google Trends could also be shown. The correlation coefficient for survey values at Welt.de is -66.06%.

At Spiegel.de this value is significantly lower at -38.91%. In contrast, the correlation coefficient for Google Trends is almost identical for both media at -28.44% for Welt.de and -28.73% for Spiegel.de.

4 DISCUSSION

In general, the analysis of the score has shown that political news consistently contains more negative than positive words. The fact that Spiegel has the lowest score and Bild the most balanced score may be related to the fact that Spiegel has the highest average text length and Bild the shortest average text length. The longer an article becomes, the more negative words can be used. The score should therefore be considered in relation to the length of the text.

The fact that AfD in particular has such a low score can have various reasons. This could be because AfD is linked to controversial issues such as the refugee crisis and Islam. However, it is also possible that the media write more negative comments and columns about AfD.

In any case, the data show that there is a correlation between a falling score of AfD articles and rising survey scores and Google trends. That a low score correlates with high Google trends could be related to the fact that scandals in particular attract a particularly high degree of attention from readers. However, this effect does not apply to the SPD and the Greens. Negative reporting correlates with these two parties not at all or only to a small measure with the survey values and Google trends.

However, it is not possible to conclude from the data whether the reporting directly influences the survey results. It is also possible that the increased survey results lead to increased reporting. A causality can therefore in no case be proven.

When analysing the share of reporting, the AfD has shown that it has remained at the same level since summer 2015. This is probably related to the refugee crisis, which began this year and is still a major topic in German politics today. As already assumed, the proportion of reporting on AfD correlates with both Google Trends and survey results. For the SPD, on the other hand, the effect is only half as strong, while for the Greens it was almost undetectable.

As only a correlation and no causality has been proven, it is not certain how the two influence each other. A higher share of reporting may be due to the increased survey values. However, it is also possible that rising survey scores lead to a higher share in reporting.

REFERENCES

ARTHUR, David ; VASSILVITSKII, Sergei: k-means++: The advantages of careful seeding / Stanford. 2006. – Forschungsbericht

spiegel.de Lkw-Fahrer rast in Menschenmenge - zwölf Tote. <https://www.spiegel.de/politik/deutschland/berlin-weihnachtsmarkt-lkw-fahrer-faehrt-in-menschenmenge-neun-tote-a-1126650.html>, . – Accessed: 2020-01-25

BIRANT, Derya ; KUT, Alp: ST-DBSCAN: An algorithm for clustering spatial-temporal data. In: *Data & Knowledge Engineering* 60 (2007), Nr. 1, S. 208–221

BUITINCK, Lars ; LOUPPE, Gilles ; BLONDEL, Mathieu ; PEDREGOSA, Fabian ; MUELLER, Andreas ; GRISEL, Olivier ; NICULAE, Vlad ; PRETTENHOFER, Peter ; GRAMFORT, Alexandre ; GROBLER, Jaques u. a.: API design for machine learning software: experiences from the scikit-learn project. In: *arXiv preprint arXiv:1309.0238* (2013)

statista.com. <https://de.statista.com/statistik/daten/studie/877238/umfrage/ranking-der-vertrauenswuerdigsten-nachrichtenquellen-in-deutschland/>, . – Accessed: 2020-01-27

stanford.edu Thomas Carlyle about the Fourth Estate. https://cs.stanford.edu/people/eroberts/cs181/projects/2010-11/Journalism/index7f0d.html?page_id=16, . – Accessed: 2020-01-27

Sigmar Gabriel verzichtet auf Kanzlerkandidatur. <https://www.spiegel.de/politik/deutschland/sigmar-gabriel-spd-chef-verzichtet-auf-kanzlerkandidatur-a-1131503.html>, . – Accessed: 2020-01-25

theguardian.com This article is more than 3 months old Germany: mass shooting attempt that killed two was antisemitic attack, minister says. <https://www.theguardian.com/world/2019/oct/09/two-people-killed-in-shooting-in-german-city-of-halle>, . – Accessed: 2020-01-25

HONNIBAL, Matthew ; MONTANI, Ines: spacy 2: Natural language understanding with bloom embeddings. In: *Convolutional Neural Networks and Incremental Parsing* (2017)

Landtagswahl 2019 in Thüringen - endgültiges Ergebnis. <https://wahlen.thueringen.de/datenbank/wahl11/wahl.asp?wahlart=LW&wJahr=2019&zeigeErg=Land>, . – Accessed: 2020-01-25

Spacy - library for advanced Natural Language Processing. <https://github.com/explosion/spaCy>, . – Accessed: 2020-01-27

RAMOS, Juan u. a.: Using tf-idf to determine word relevance in document queries. In: *Proceedings of the first instructional conference on machine learning* Bd. 242 Piscataway, NJ, 2003, S. 133–142

statista.com. <https://de.statista.com/statistik/daten/studie/154154/umfrage/anzahl-der-visits-von-nachrichtenportalen/>, . – Accessed: 2020-02-13

REMUS, Robert ; QUASTHOFF, Uwe ; HEYER, Gerhard: SentiWS-A Publicly Available German-language Resource for Sentiment Analysis. In: *LREC* Citeseer, 2010

Die große Schulz-Show. <https://www.spiegel.de/politik/deutschland/martin-schulz-erste-rede-des-spd-kanzlerkandidaten-a-1132226.html>, . – Accessed: 2020-01-25

Nichtwähler verhelfen SPD zu Umfragehoch. <https://www.spiegel.de/politik/deutschland/schulz-effekt-nichtwaehler-verantwortlich-fuer-hoch-der-spd-a-1134258.html>, . – Accessed: 2020-01-25

SPD zählt über 10.000 Neu-Mitglieder seit Schulz-Nominierung. <https://www.waz.de/politik/spd-zaeht-ueber-10-000-neu-mitglieder-seit-schulz-nominierung-id209802495.html>, . – Accessed: 2020-01-25

spiegel.de 20 Jahre SPIEGEL ONLINE. <https://www.spiegel.de/netzwelt/web/spiegel-online-feiert-20-geburtstag-a-999169.html>, . – Accessed: 2020-01-25

STEINLEY, Douglas: Validating clusters with the lower bound for sum-of-squares error. In: *Psychometrika* 72 (2007), Nr. 1, S. 93–106

VAN DER MAATEN, Laurens: Accelerating t-SNE using tree-based algorithms. In: *The Journal of Machine Learning Research* 15 (2014), Nr. 1, S. 3221–3245

APPENDIX

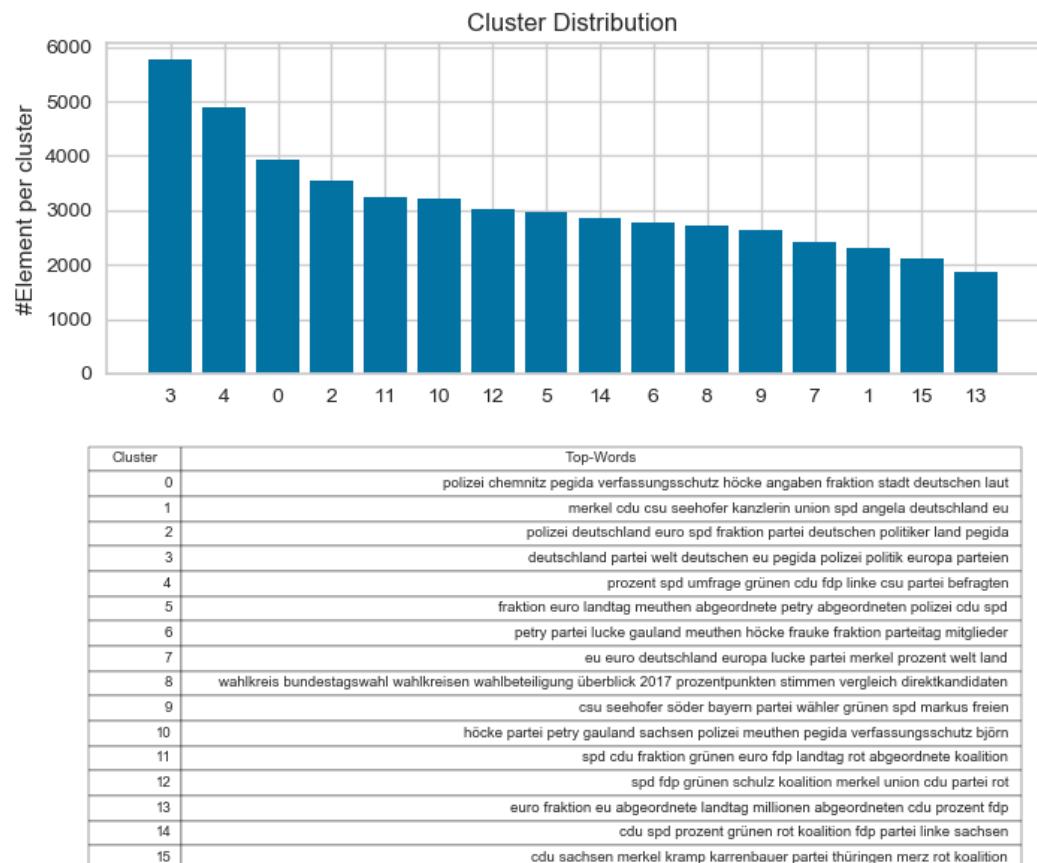


Figure 24: Cluster Distribution of AfD articles and Top-Ten-Words

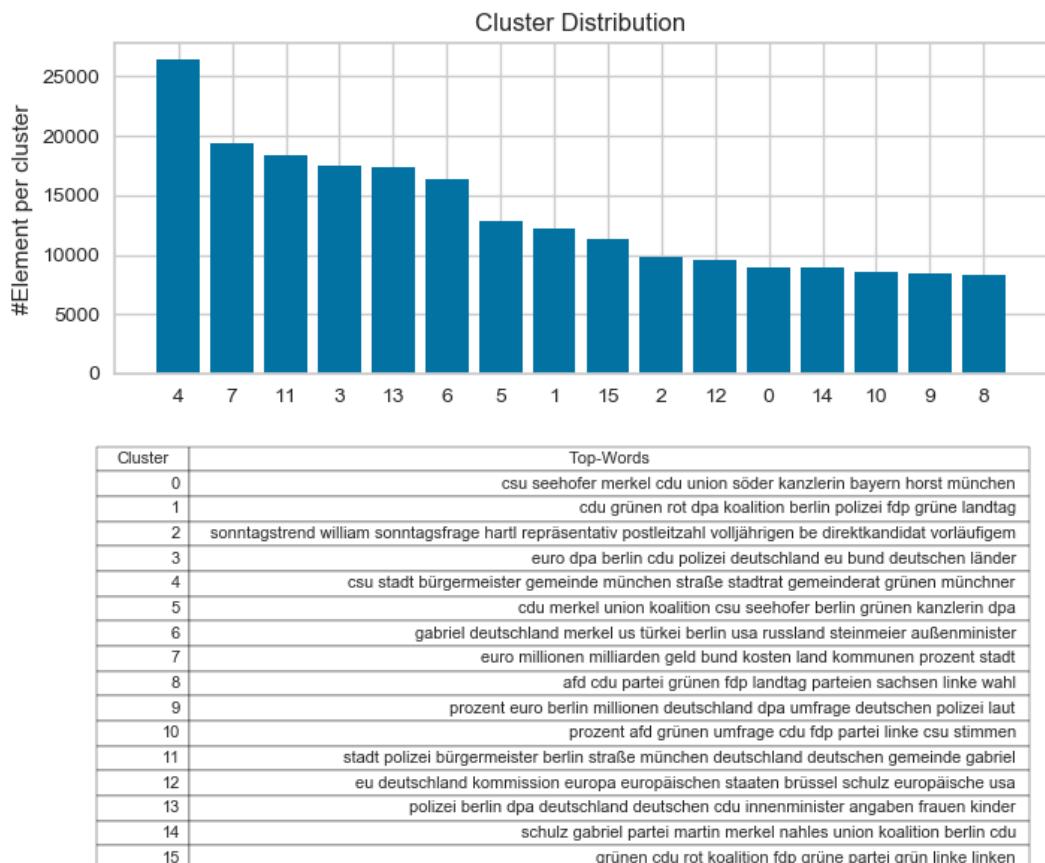


Figure 25: TSNE Projection of all SPD articles

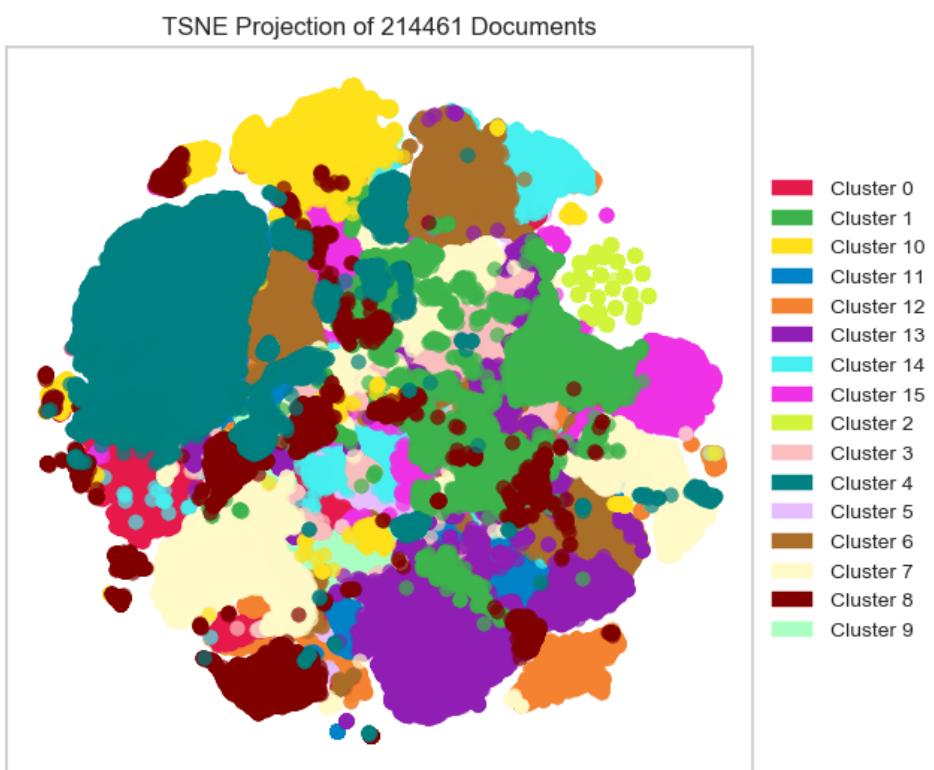


Figure 26: TSNE Projection of all SPD articles

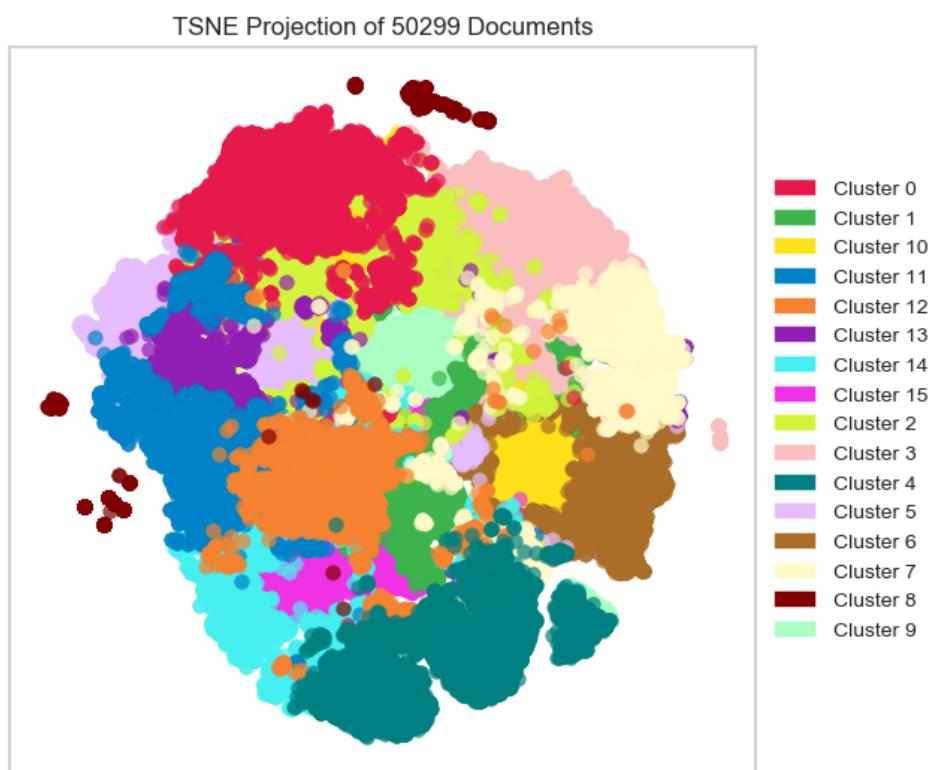


Figure 27: TSNE Projection of all AfD articles

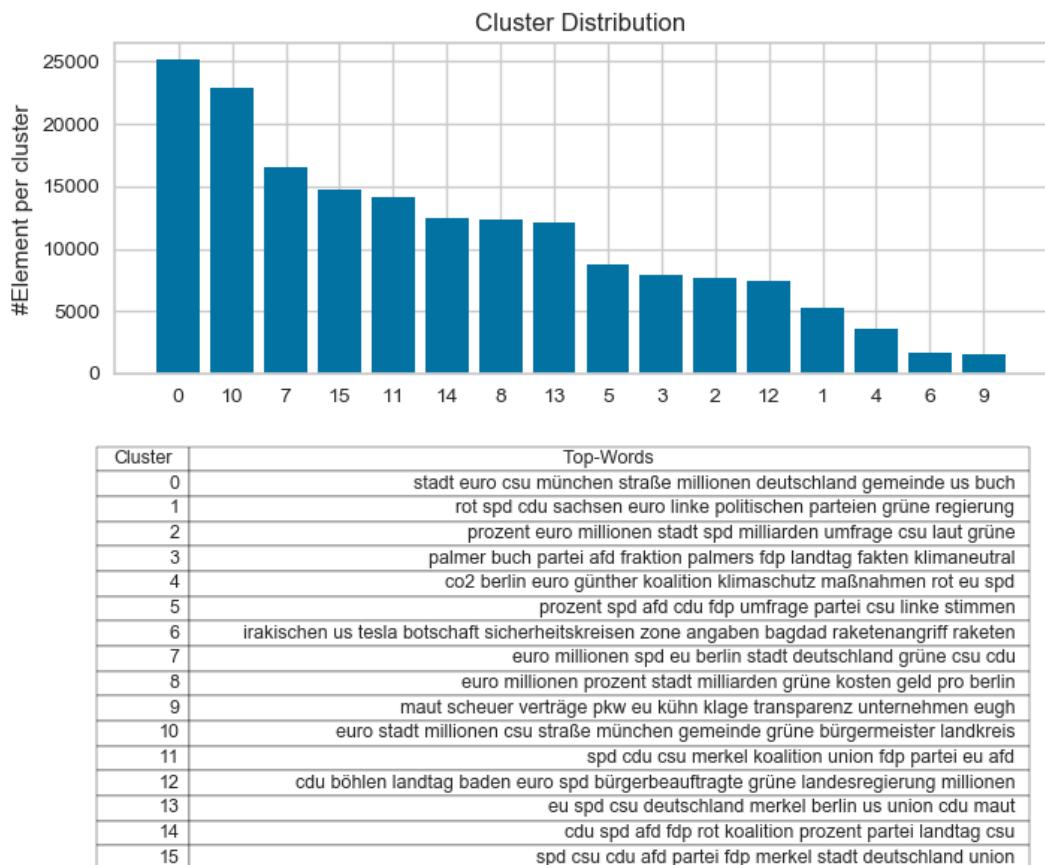


Figure 28: TSNE Projection of all Green articles

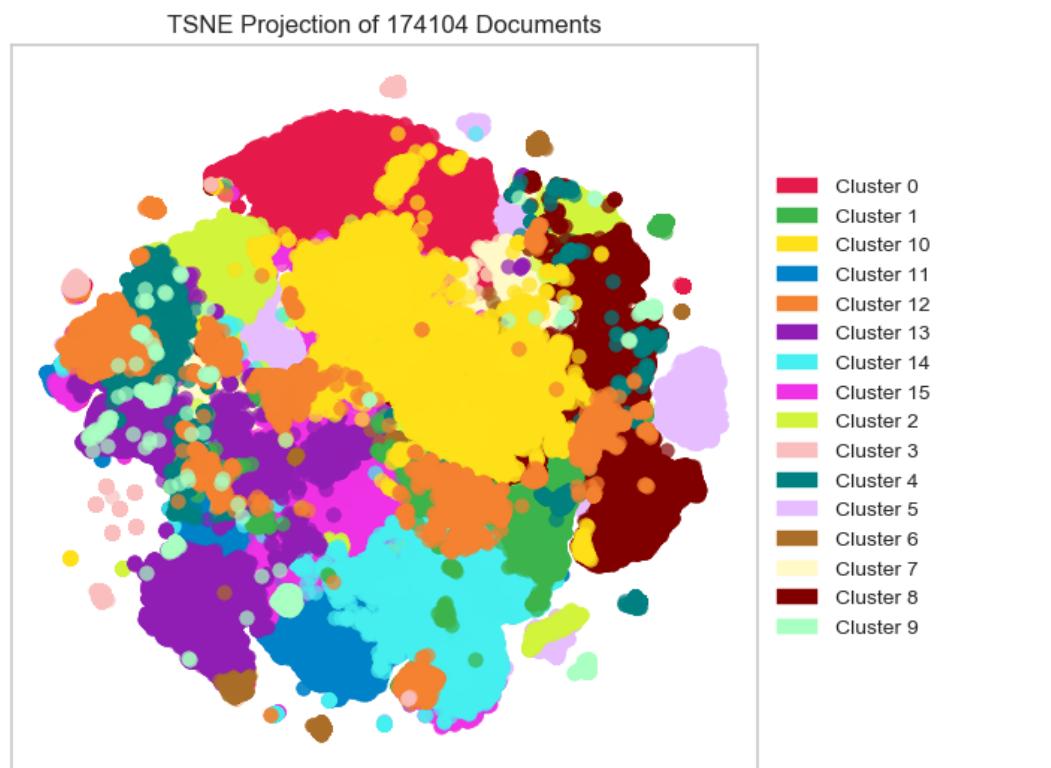


Figure 29: TSNE Projection of all Green articles

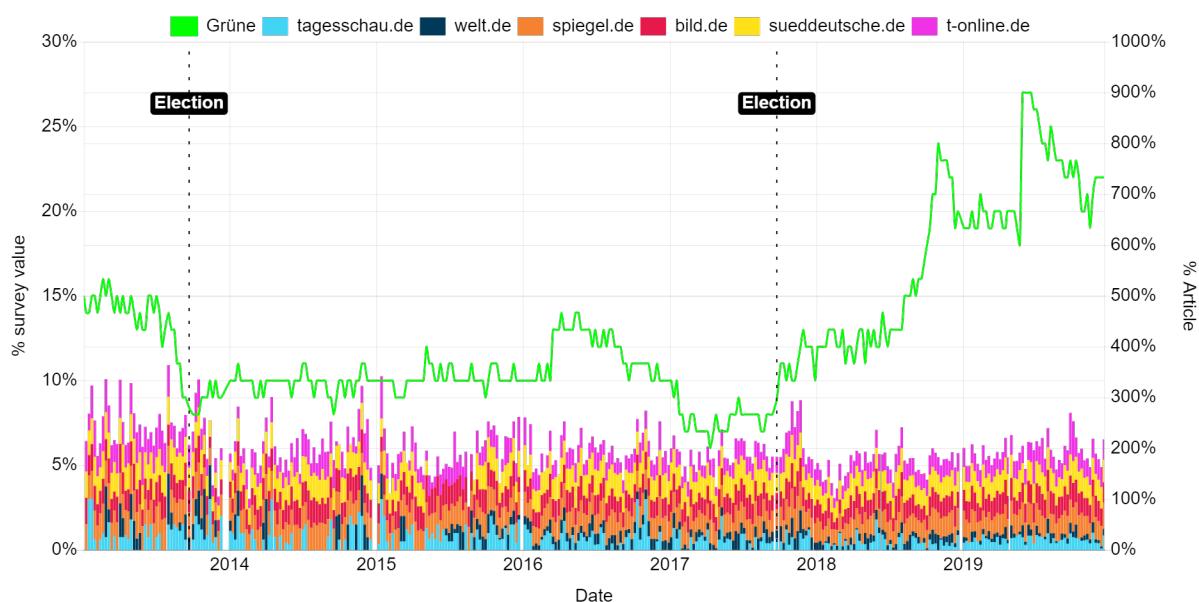


Figure 30: Percentage of reporting about the Greens compared to the survey values

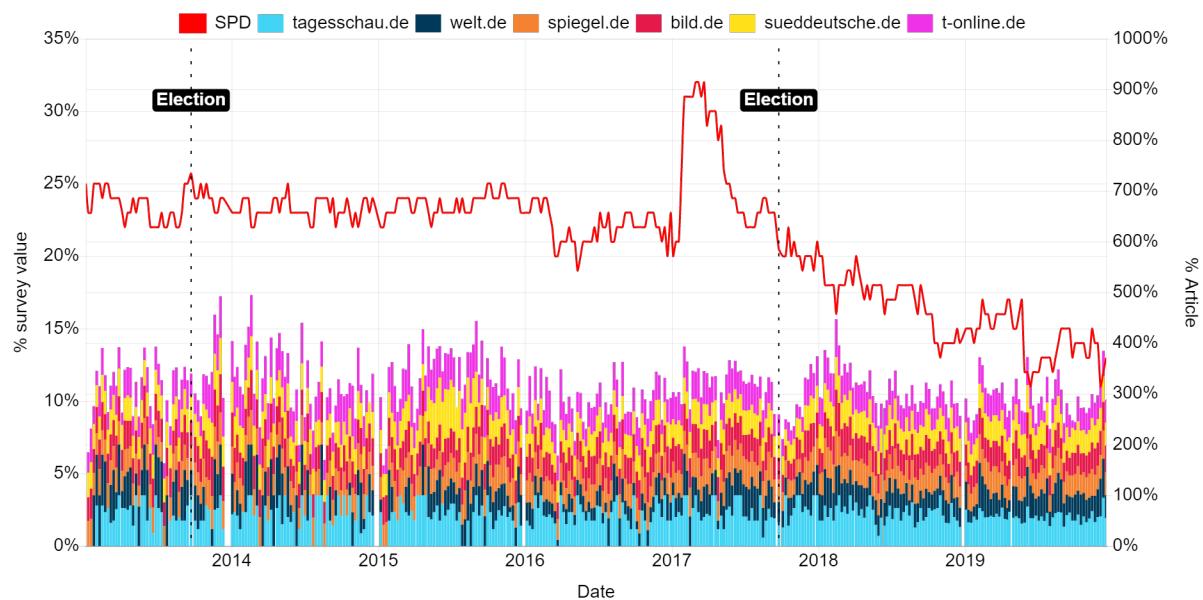


Figure 31: Percentage of reporting about the SPD compared to the survey values

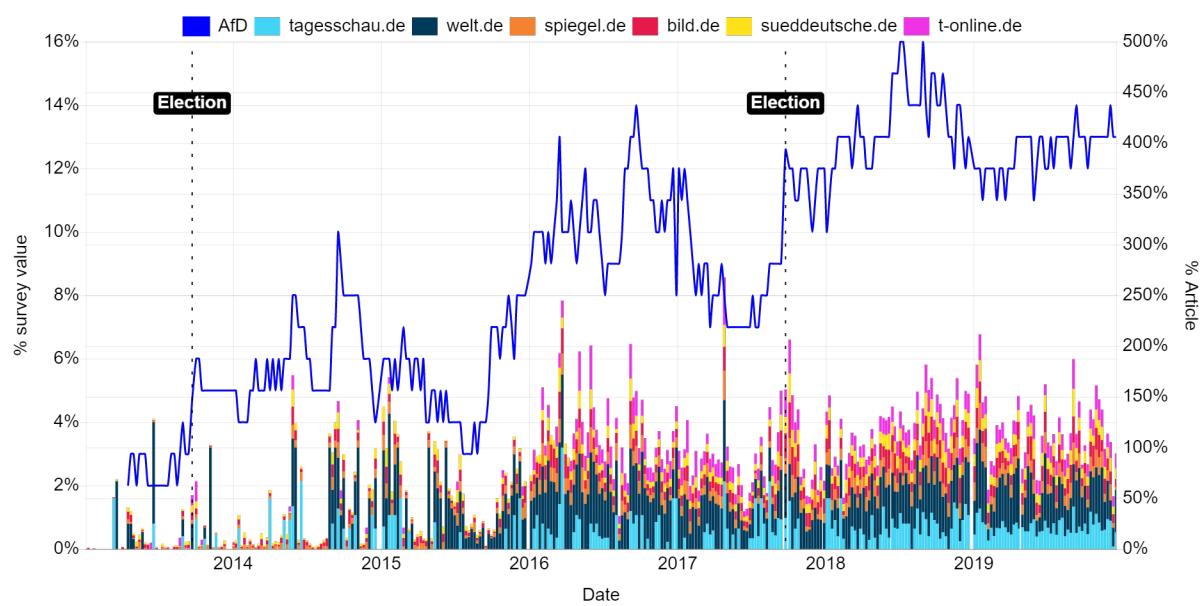


Figure 32: Percentage of reporting about the AfD compared to the survey values

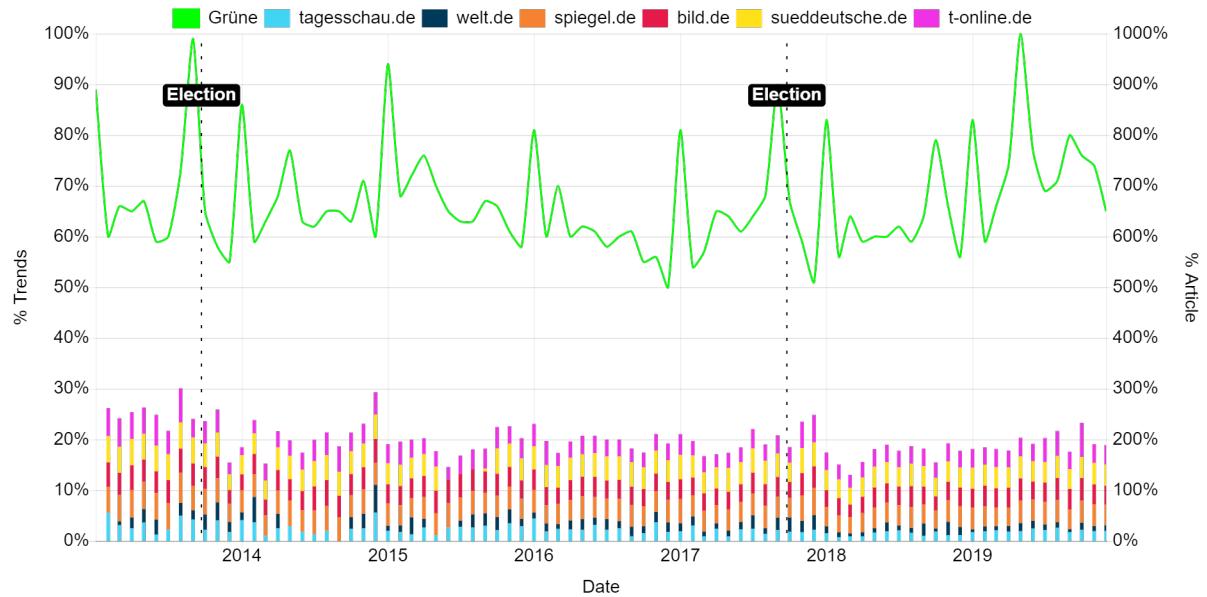


Figure 33: Percentage of reporting about the Greens compared to Google Trends

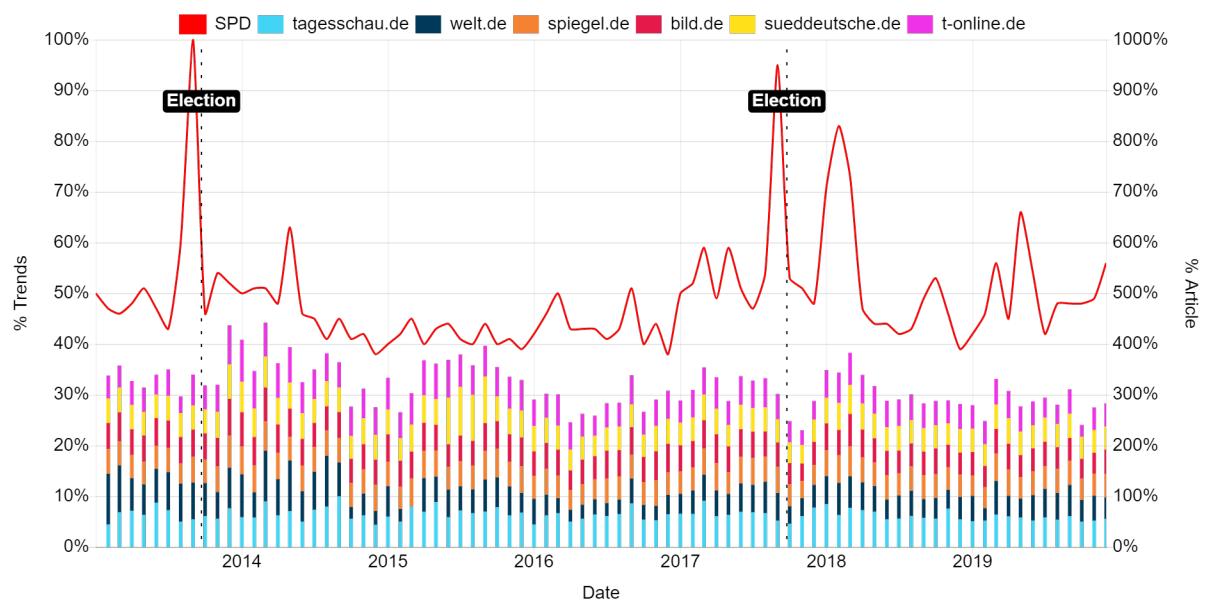


Figure 34: Percentage of reporting about the SPD compared to Google Trends

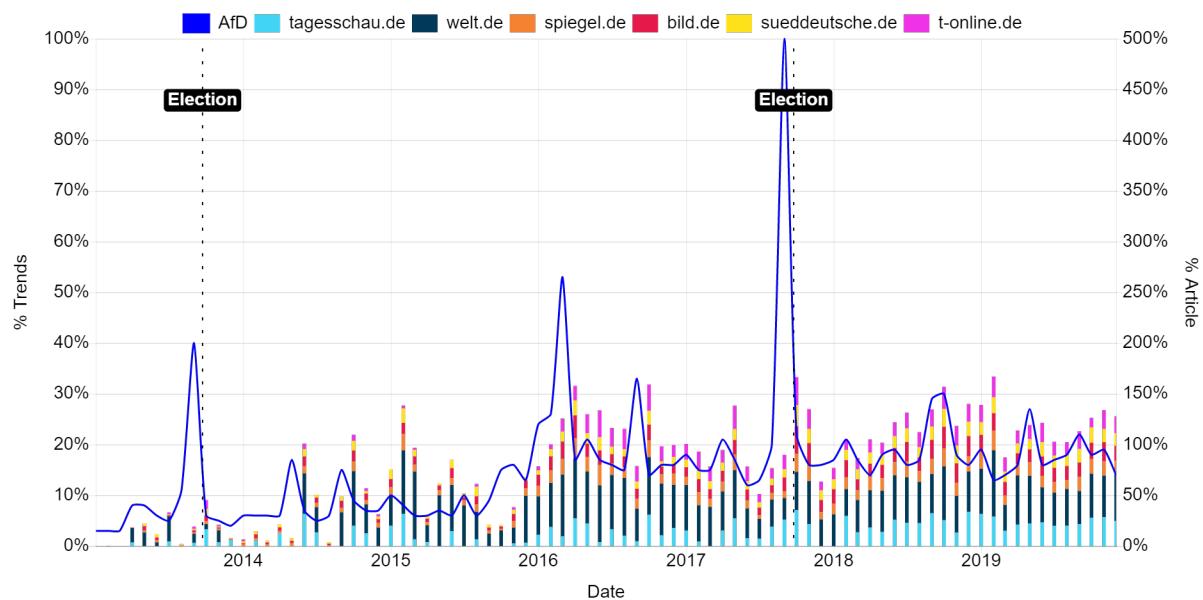


Figure 35: Percentage of reporting about the AfD compared to Google Trends

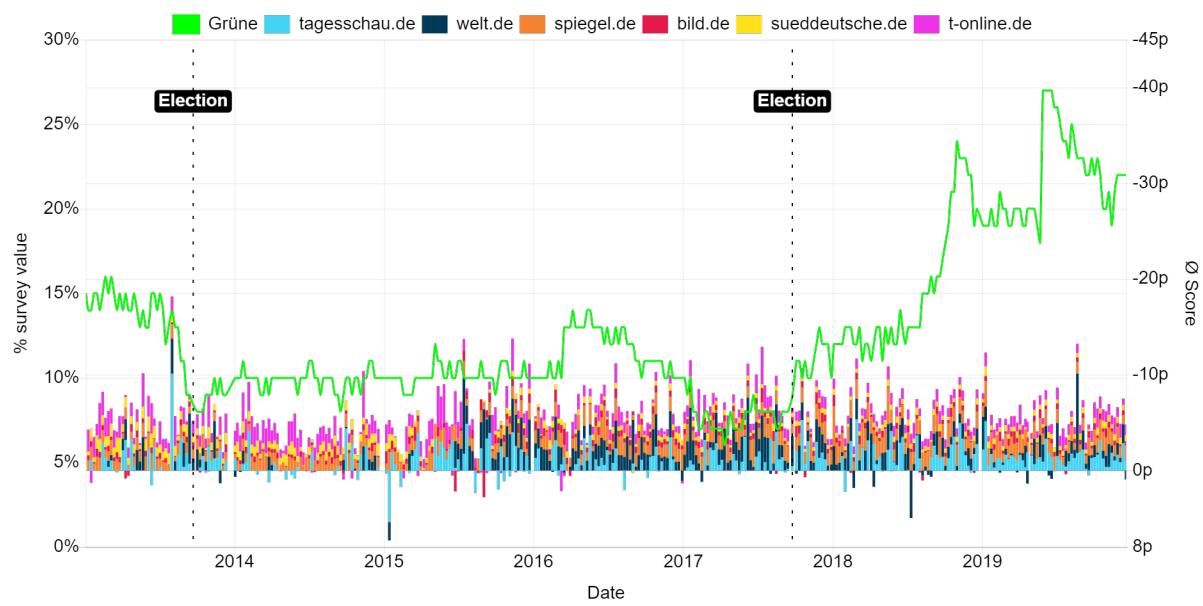


Figure 36: Average score about the Greens compared to the survey values

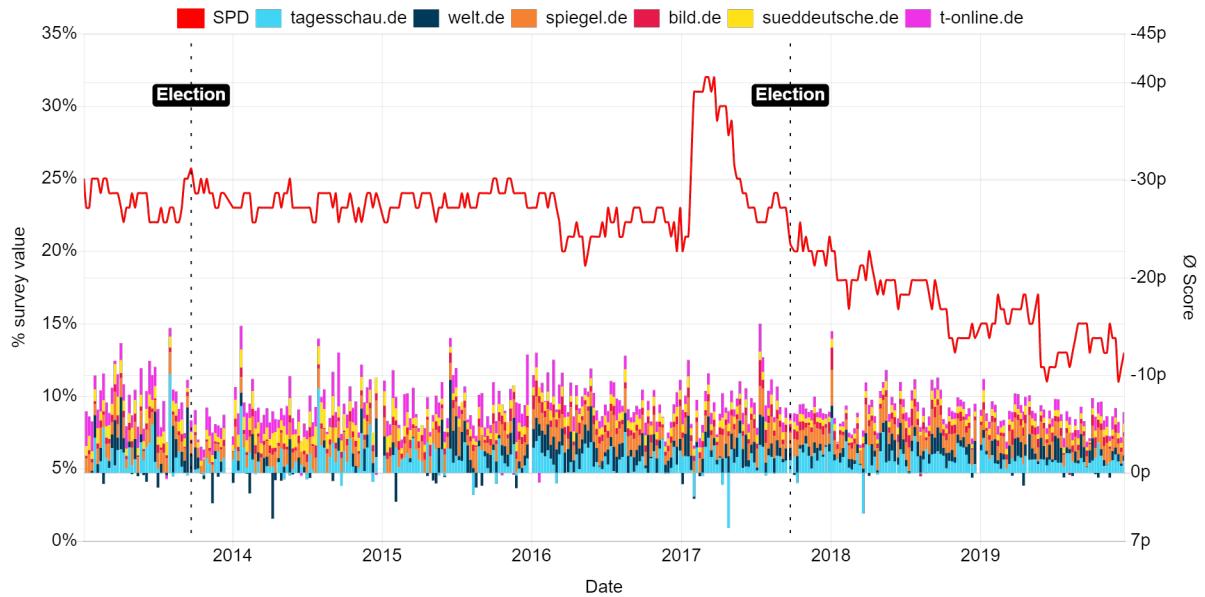


Figure 37: Average score about the SPD compared to the survey values

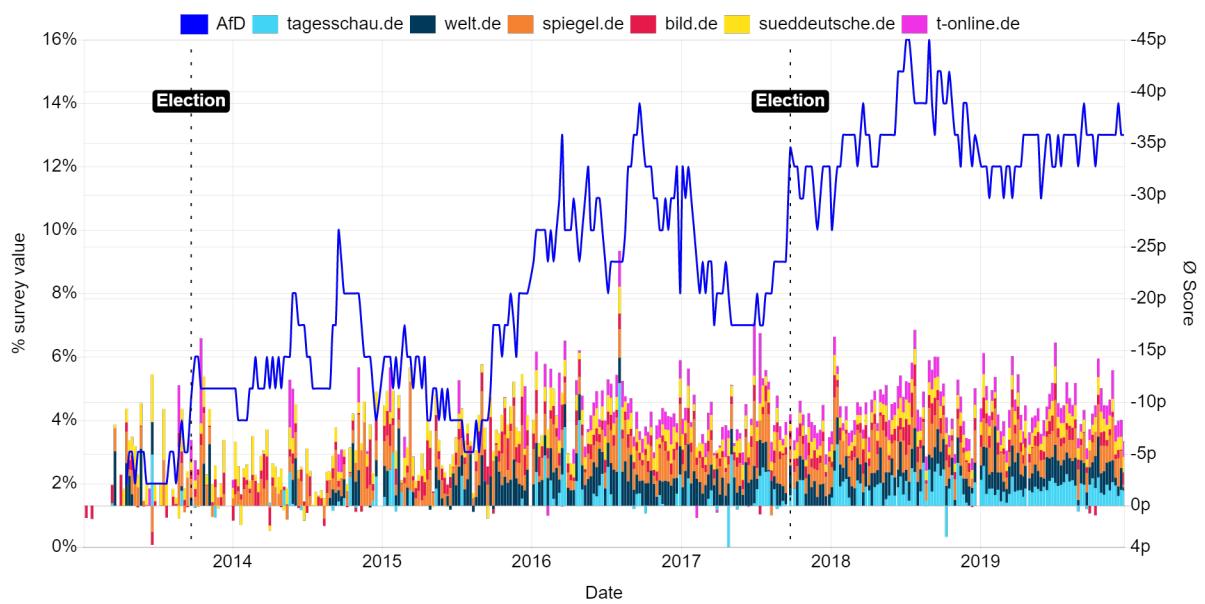


Figure 38: Average score about the AfD compared to the survey values

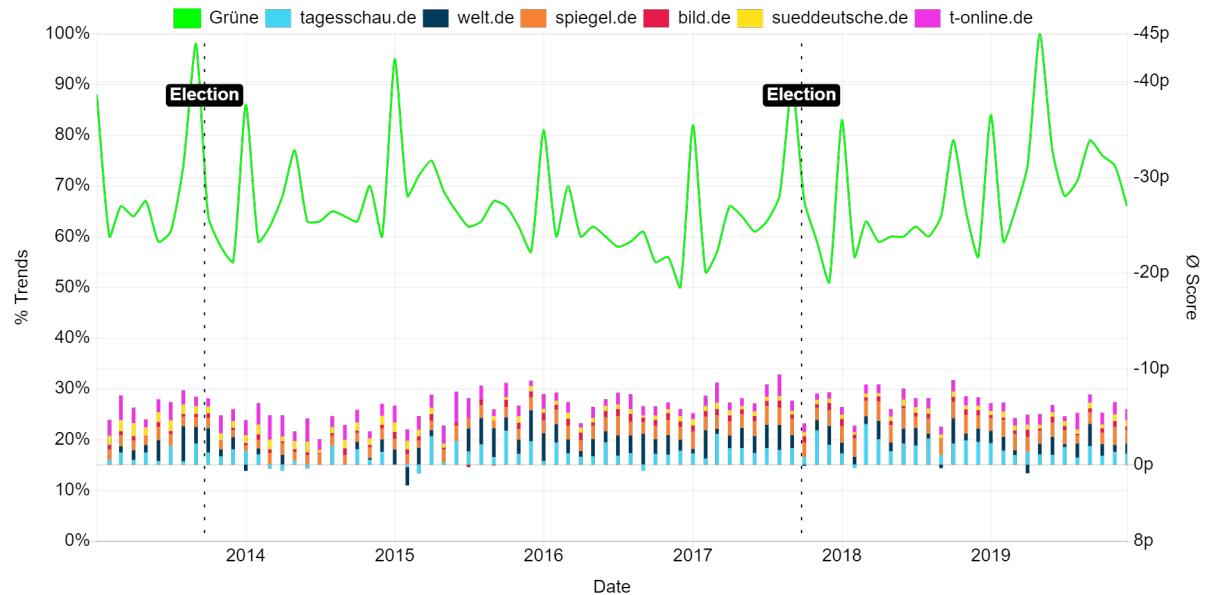


Figure 39: Average score about the Greens compared to Google Trends

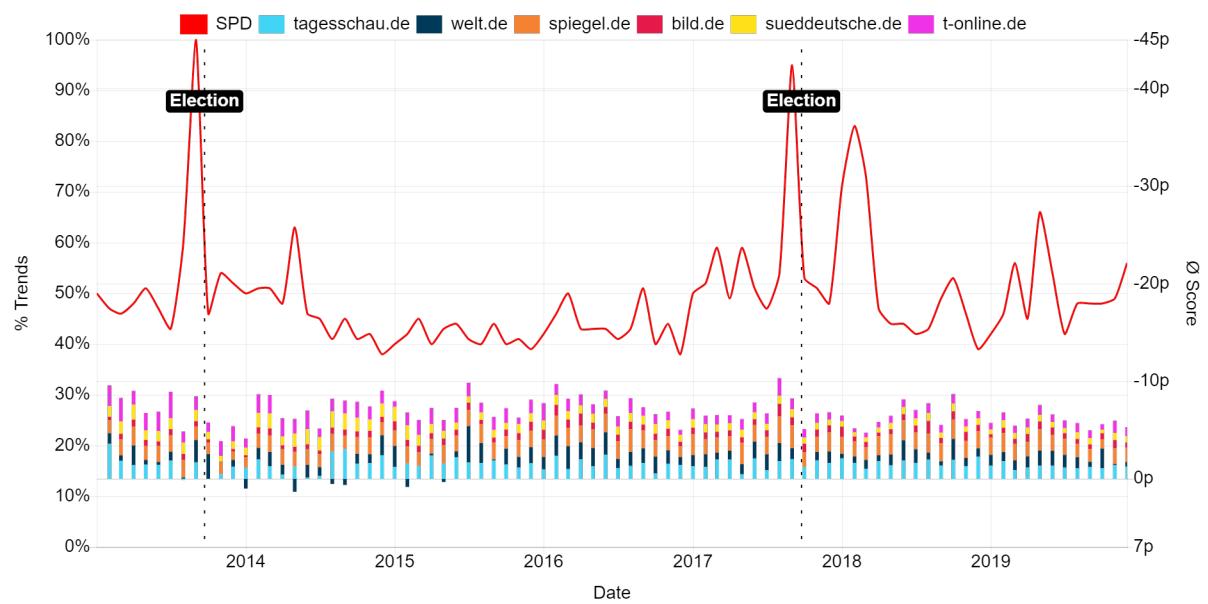


Figure 40: Average score about the SPD compared to Google Trends

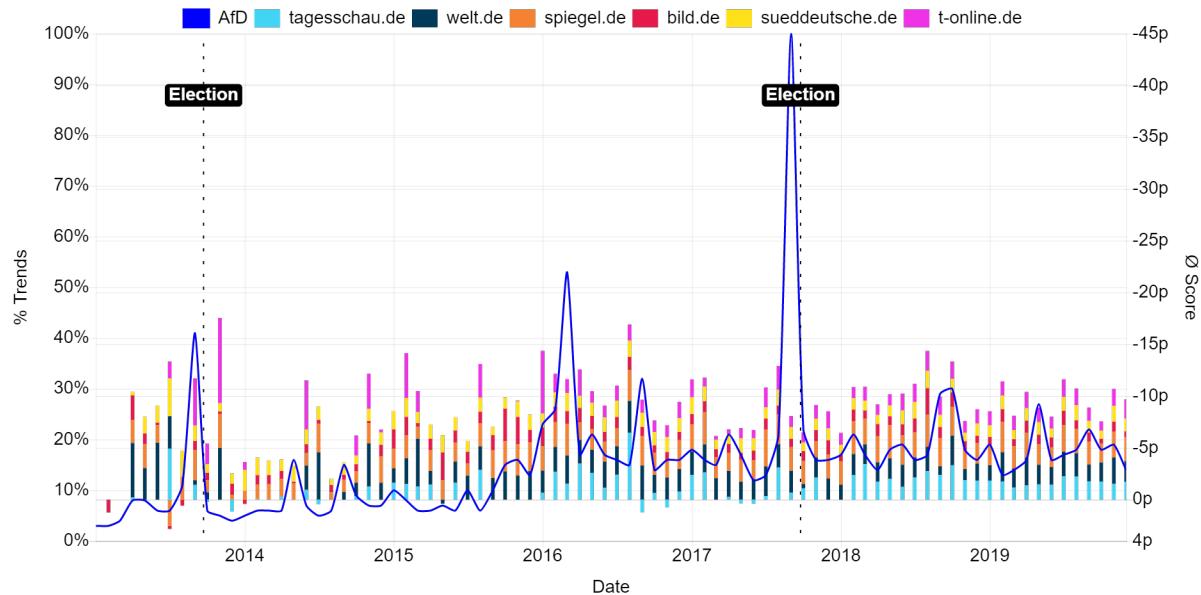


Figure 41: Average score about the AfD compared to Google Trends

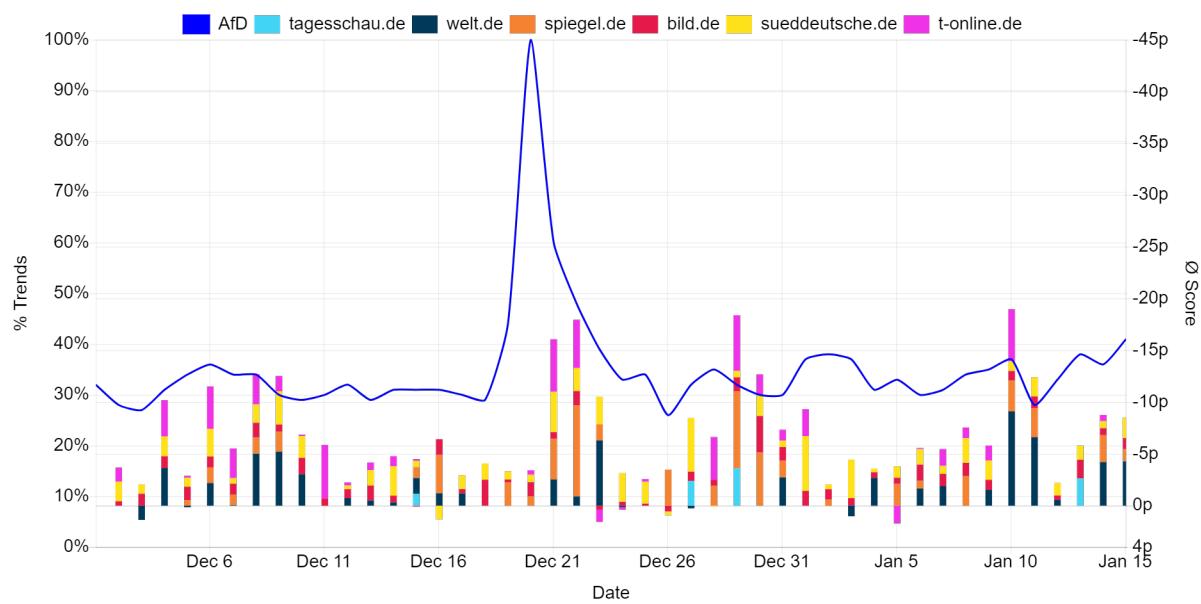


Figure 42: Average score about the AfD compared to Google Trends in the period from 01.12.16 to 15.01.17

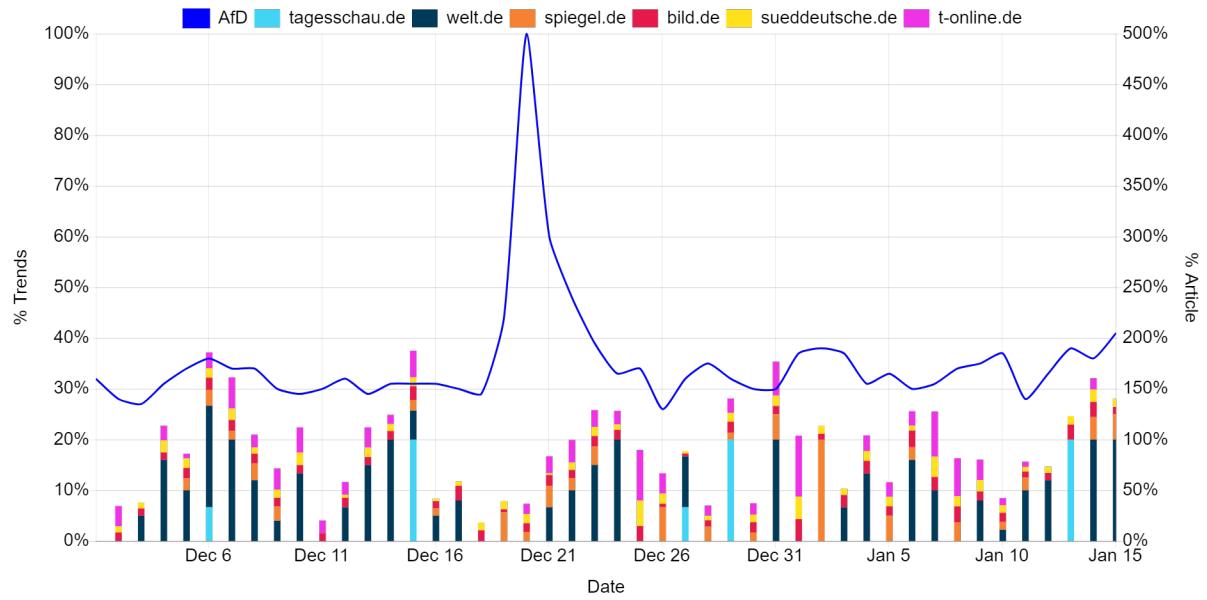


Figure 43: Percentage of reporting about the AfD compared to Google Trends in the period from 01.12.16 to 15.01.17

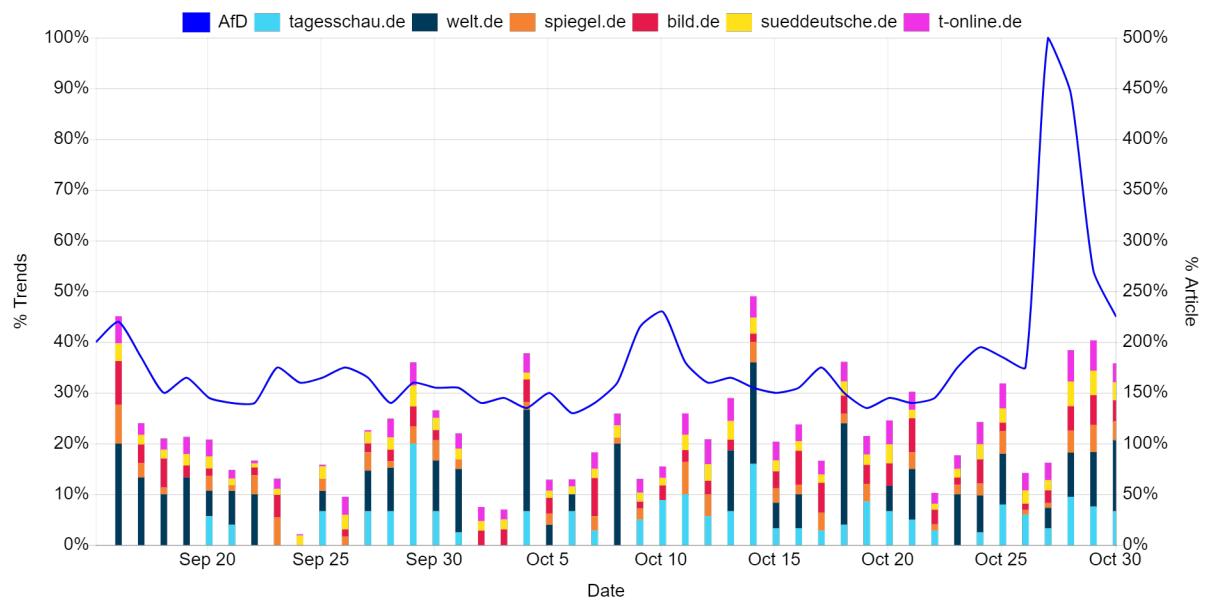


Figure 44: Percentage of reporting about the AfD compared to Google Trends in the period from 15.09.19 to 29.09.19

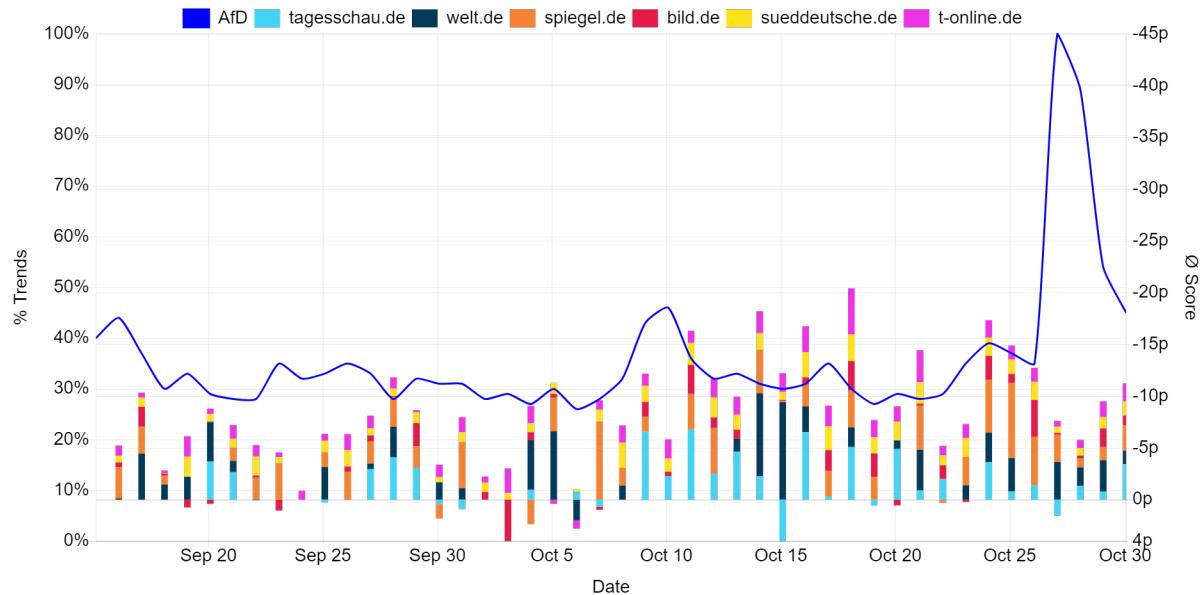


Figure 45: Average score about the AfD compared to Google Trends in the period from 15.09.19 to 29.09.19

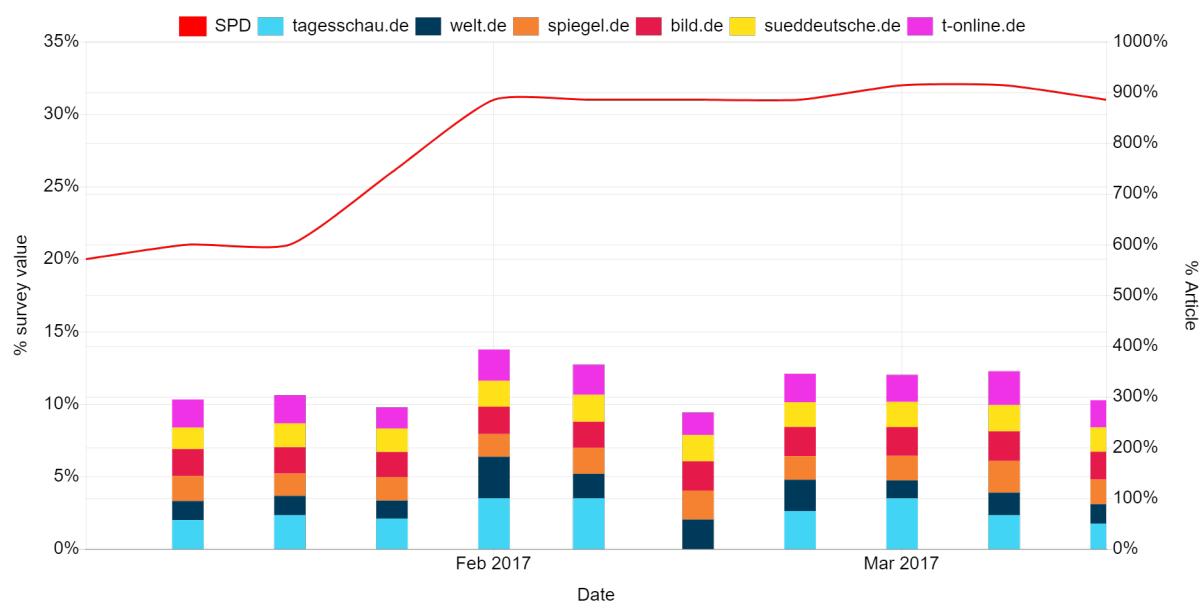


Figure 46: Percentage of reporting about the SPD compared to Google Trends in the period from 01.01.17 to 15.03.17

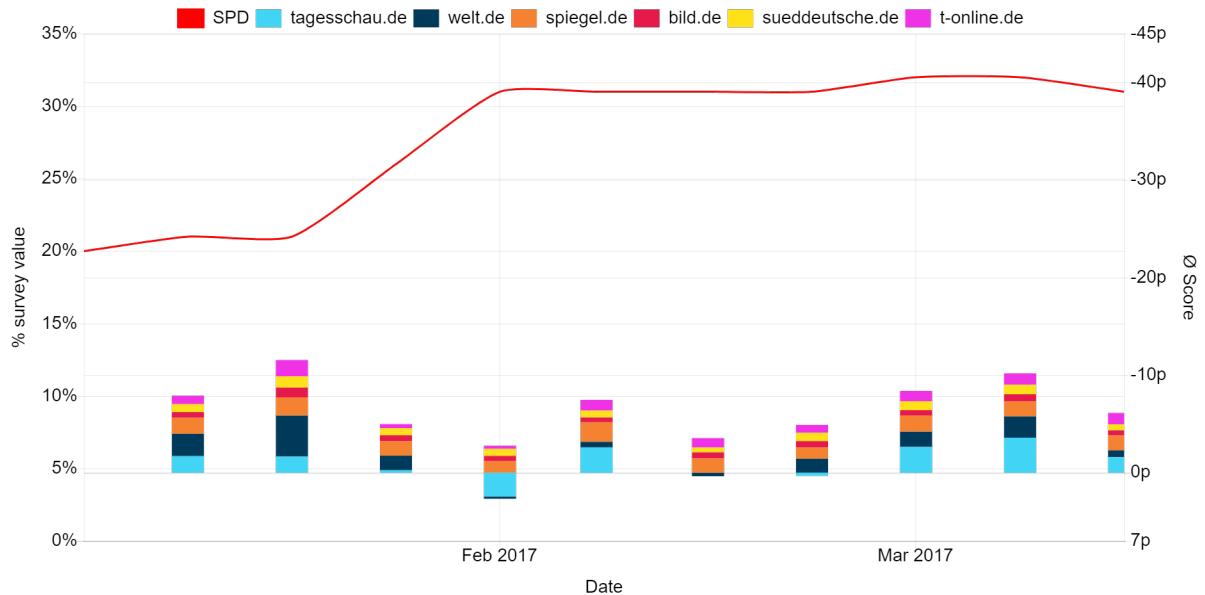


Figure 47: Average score about the SPD compared to Google Trends in the period from 01.01.17 to 15.03.17

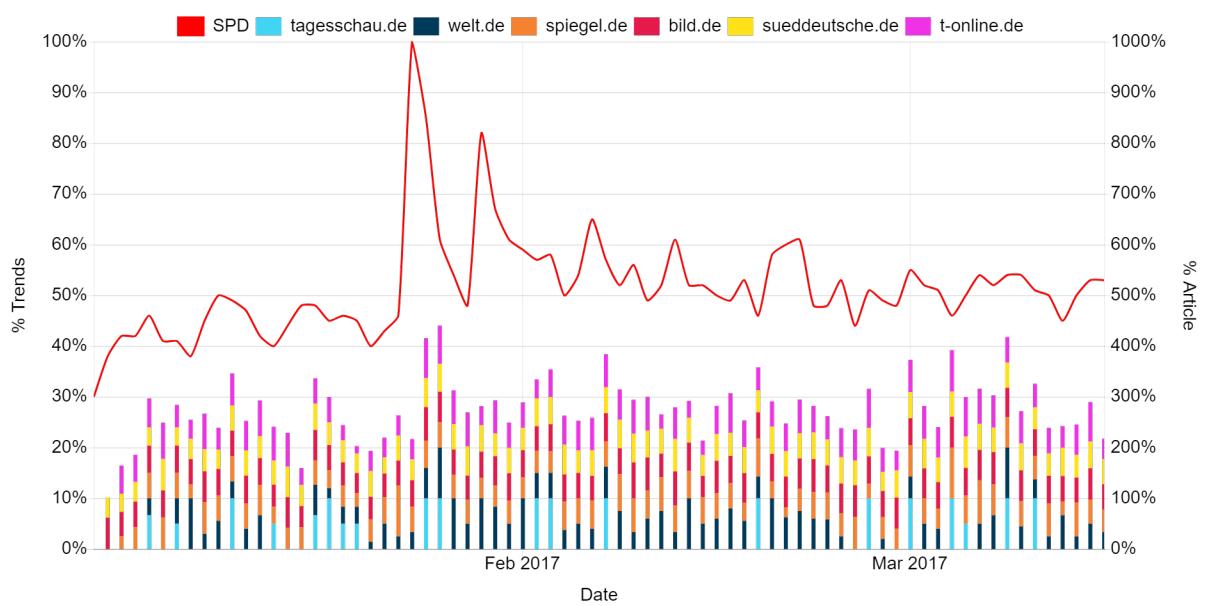


Figure 48: Percentage of reporting about the SPD compared to Google Trends in the period from 01.01.17 to 15.03.17

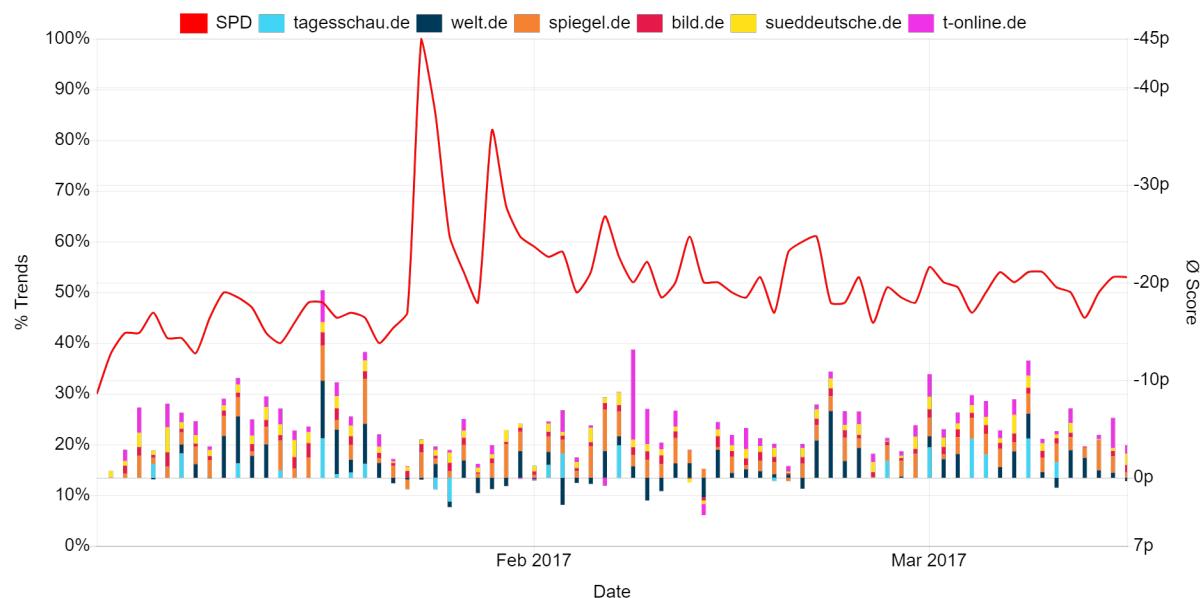


Figure 49: Average score about the SPD compared to Google Trends in the period from 01.01.17 to 15.03.17

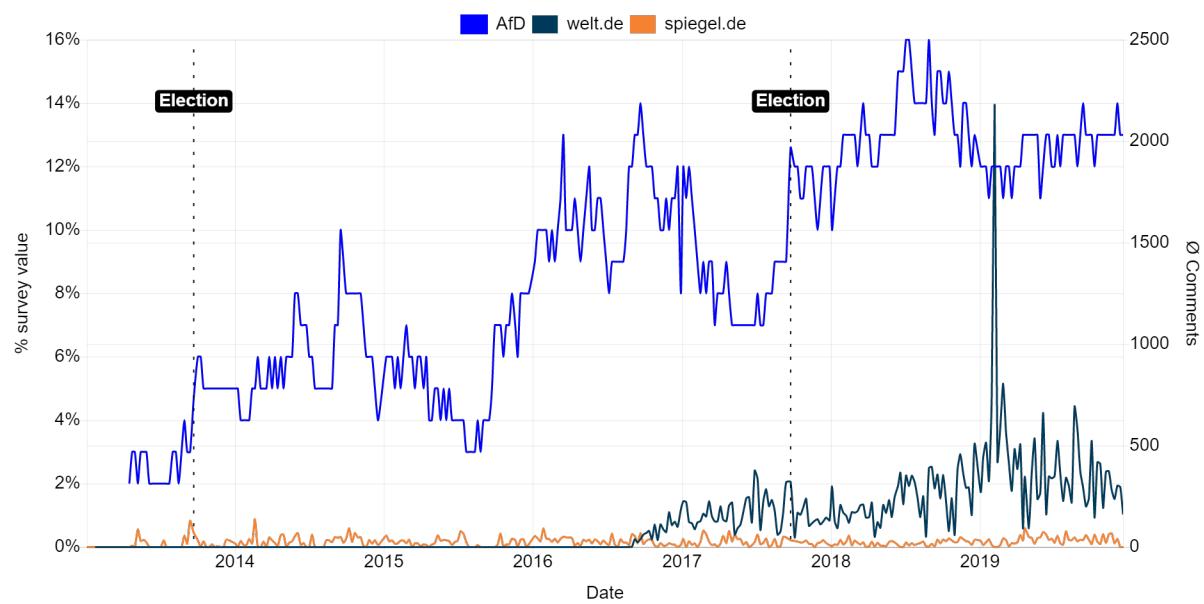


Figure 50: Average amount of comments compared to the survey values

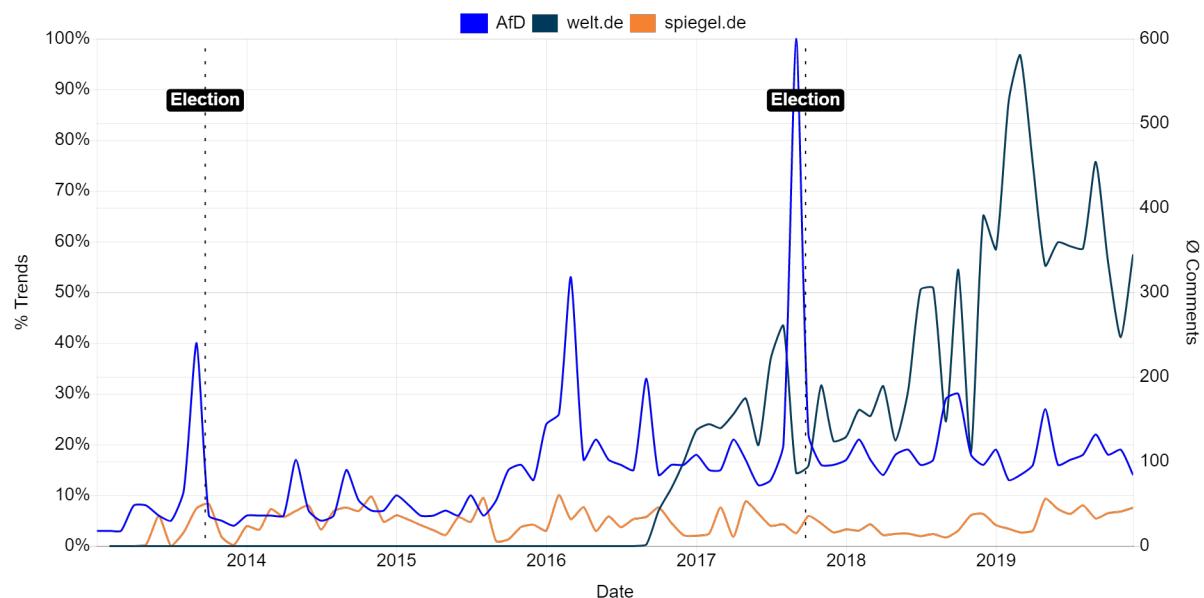


Figure 51: Average amount of comments compared Google Trends

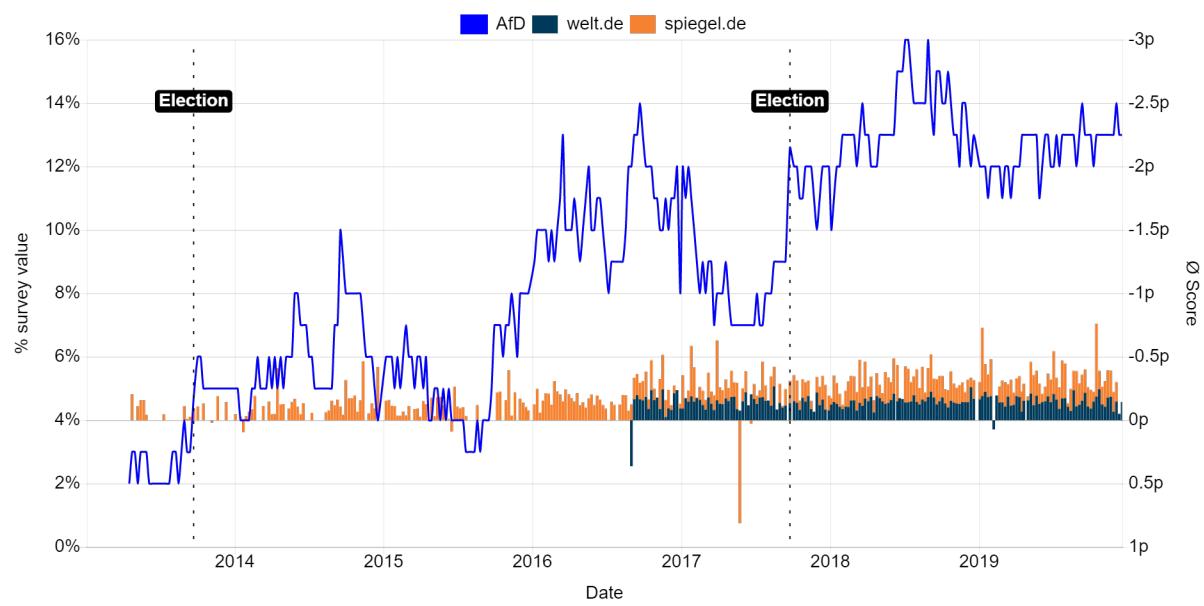


Figure 52: Average score of comments compared to the survey values

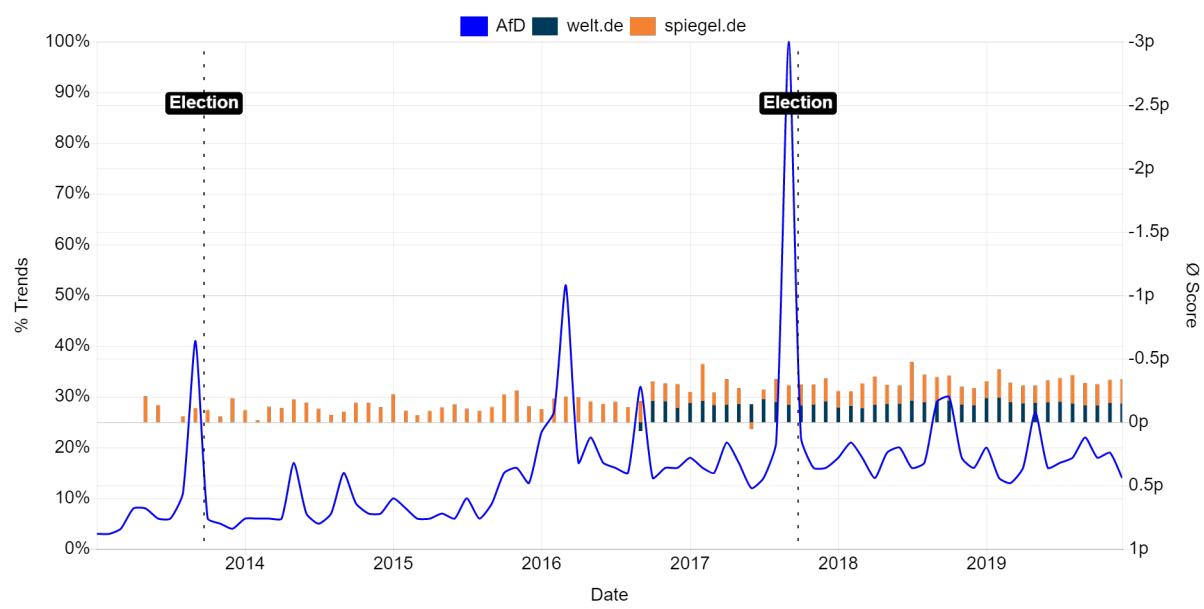


Figure 53: Average score of comments compared to Google Trends