THE CHANGES IN THE ALTERNATIVE FÜR DEUTSCHLAND AGENDA SETTING*:*

An analysis of how the newspapers portray these changes in their articles

Introduction

The *Alternative für Deutschland* (AfD) – Alternative for Germany, in English – party is a political phenomenon. It emerged in 2013 as a single issue party, identified as right-wing and populist, against the euro. A few months later they almost achieved the threshold to enter in the German Bundestag. One year after the party got 7 seats in the European Parliament. It shifted its agenda, became more radical and focused on the immigration crisis after a dramatic change in its leadership. In 2017 AfD had 12,6% of the votes in the Bundestag elections and became the third largest party in the Bundestag (Berning, 2017; Goerres et al., 2018; Grimm, 2017).

The trajectory of AfD became an object for study by political scientists. It is a unique phenomenon, because it was the most successful right-wing party since the end of the World War II in Germany and one with the best results in the currently European politics (Goerres et al., 2018). The presence of the media also helped the party's increase in influence. The headlines involving the party gave to the population an alternative solution for the problems in Germany and to the crises. In this sense, people desperate for keeping their jobs and to stabilize the economy, concerned with the measures assumed by Merkel during the crisis, gave their vote to AfD. Thus, the newspapers were decisive to picture the AfD as a solution for those not in favor of the paths Germany was taking through the past years (Chazan, 2017; Knight, 2018).

This article aims at checking the relation between the shifts in the history of the party and the articles in the media. The goal is to answer**:**

Q1: What is the overall portray of the party to the newspapers' readers? What do they most read about AfD?

Q2: Regarding to the topics, is a time clustering an efficient way to demonstrate the shifts in the agenda of the party through the news? Are there any other way of cluster this articles?

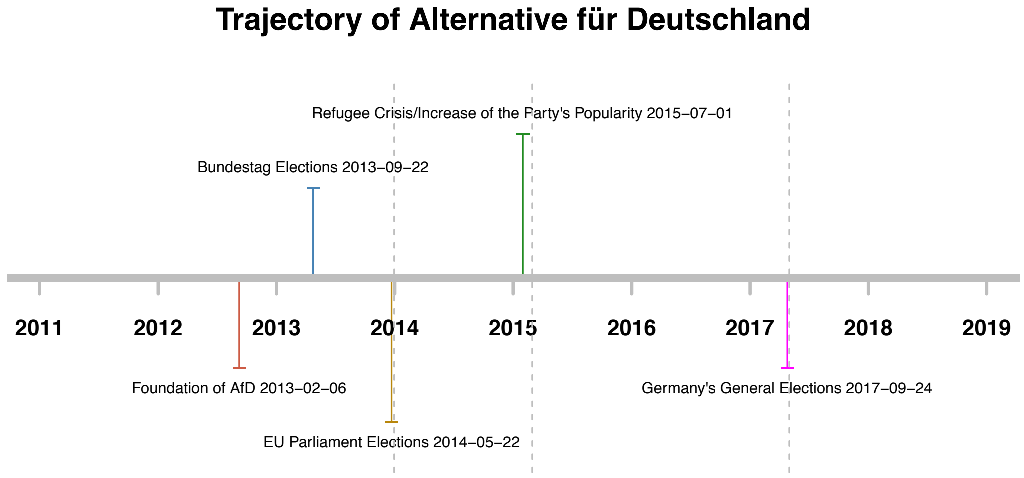
I will use analytical mechanisms in the R programme. I will evaluate the articles written about the party during the almost six years of its existence. In this sense, I divided my data into four phases, considering the main events related to AfD. I will compare these phases selected by the historical background to clustering process using the most typed words in the articles. The articles used were from The Guardian and Deutsche Welle newspapers. The first will present an international perspective, while the second will portray how the party is referred in the domestic media.

In the first section I will briefly describe the historical trajectory of the AfD. I will present a historical timeline and will explain the four phases. In the second section I will describe the data collected and explain the reasons under the choice of the two newspapers. The third section I will describe the tools and features used for the analysis. The last section will bring the results from the analysis.

1 The Historical Trajectory of The Alternative für Deutschland

The *Alternative für Deutschland* emerged in February 2013, in Germany, as a single-issue party, anti-Euro but not against the EU. Some classify the party as a right-wing populist party due to its agenda opposing to the economic integration of the EU. Besides, the line adopted on anti-immigration policies and the nationalistic features in the manifestos reaffirm their eurosceptic and right-wing populist design (Schmitt-Beck, 2017).

Its history involves party-struggles and significative changes in the agenda setting. This is decisive to understand how the party is portrayed in the media and how it gained success in the German political scenario. It is divided into five major events and four phases. In the timeline, it is possible to observe these events and the phases encompassing the development of the party. The phases are split by the gray lines.



Founded by conservatives unhappy with the centrist directions of Merkel, the AfD used the crises in the European Union to establish themselves. In 2013, before the German General Elections, their agenda was against the Eurozone. For them, the zone was underestimating the power of Germany. Besides, they showed concern towards the national budget and responsibilities assumed by the government on the eurozone crisis. In the same year, they almost achieved the minimum threshold of 5% to enter in the Bundestag. With 4,7% of the votes, it was the first eurosceptic party to attract substantial electoral support in such a short period of time (Berning, 2017; Grimm, 2017; Lees, 2018).

Until the European Parliament elections in 2014, the AfD gained more space in the political scenario. The party was known by opposite suggestions regarding measures adopted by the government. However, in 2014, the party struggles and internal conflicts started. In that year, the party achieved 7,1% of the votes for the EU parliament, which represented seven seats. This was decisive for the establishment of the party as an right wing opposition. At this point, they gained visibility in the international media by influencing other right wing parties (Knight, 2018; Schmitt-Beck, 2017).

After the elections in 2015, Bernd Lucke, one of the founders of the AfD left the party and created the Alliance for the Progress party. By one hand, some leaders, including Lucke, were moderate market-liberal eurosceptics. By the other, it was increasing the leadership supporting a more national conservative scope, focused on the migration issue. These struggles changed the agenda setting of the party after the exit of Lucke. They gained more nativist roots, against the measures assumed by Merkel towards the migratory crisis. Consequently, after the summer of 2015, where Germany received a significant amount of refugees, they changed its scope of agenda and focused more on migration. Hence, the eurozone situation opened the grounds for the party, but it started to gain more political importance when shifted to the migratory issue (Art, 2013; Grimm, 2017; Lees, 2018).

In 2017, the Bundestag Elections marked the party in history. It was the first right-wing party, since 1949, to gain seats in the parliament and to reduce the seats of the most established party, the Christian Democratic Union with Merkel as leader. With 12,6% of the votes, and a mobilization of a lot of non-voters, the party became the third largest party in terms of seats. Later, the AfD gained more attention in the media, more followers and became an important phenomenon to be studied (Berning, 2017; Lees, 2018).

Arzheimer (2015) elaborated an study about the most quoted topics in the party manifestos for the Bundestag Elections in 2013. The most-related themes to AfD were biased on monetary issues. Lees (2018) argued about the voters' profile for the 2017 elections. Aa dramatic change happened: the mobilization of more than one million of voters that didn't attend in previous elections – people concerned with their jobs and savings due to the migration crisis. Hence, it is clear a shift in the agenda and in the voting behavior of the German electorate.

The historical trajectory of the AfD is translated into four phases. The first phase until the European Parliament elections, the second until the trigger of the migration crisis, a third between the crisis and the elections in 2017 and the last one, the fourth, after this period. These four phases are marked by specific events that shaped not only the political agenda of the AfD, but also the opinion and behavior of the voters. I will use these phases to see if the media accompanied this shifts in their articles of it is portrays the party in one specific way.

2 Data collection

This article will analyze two newspapers: The Guardian and Deutsche Welle (DW). Both are in English, to facilitate the topic analysis and to guarantee that the results obtained from the scraping and machine learning are accurate.

During the newspaper's selection, I aimed at working with two different scopes: one international and one local. My goal is to see if there are different patterns in the articles when dedicated to the domestic or to the international public. It is visible a shift in the agenda of the AfD. Here, I want to observe if there are any shifts in the articles' topics.

The Guardian is an international well known newspaper, with daily articles. The newspaper has specific sections for continents and regions, and some specific countries.

Germany has its own section in the newspaper, increasing the articles' focus. Due to the relevance of the newspaper, the written language and the specific section about Germany, I adopted it to this research.

DW newspaper presents an interesting scope. The newspaper is written in German and in English, which means that almost all articles are translated to the English language, providing a domestic perspective of a newspaper in an international language. My focus was having two newspapers written in the same language, to avoid bias or losing more technical and theoretical terms when analyzing in other languages. The DW newspaper is one of the biggest in the country, it tends to report the more important facts. This helps by guaranteeing that there will be articles written about politics and elections and, probably, about the party I am writing about.

The scrapping process was done during November month and finalized on 01.12.2018. The time frame from the collected data is from 01.01.2013 to 30.11.2018. I will analyze the entire period from the creation of the party to the day before I finished the scrapping process. It was expected, from the beginning, a less amount of articles from The Guardian compared to DW. This is because one is an international newspaper that do not focus only in German news. However, it is important for the analysis two papers with different scopes to observe their topics through the years and the way they portray the party.

a. The Data Collection in The Guardian newspaper: scope and challenges

The articles' collection in The Guardian newspaper was through the R programme. This data collection was made by scrapping the pages related to Germany in the newspaper. The Guardian newspaper has a specific page for news about Germany and the access is by the main label "world" and, after, "Germany".

I scraped the first 265 pages of the "Germany" section. It referred to the data setting from 01.01.2013 to 30.11.2018, resulting in 5273 articles published about Germany during almost six years. However, I encountered some problems while scraping. First, all titles were duplicated. To remove I used a *gsub* function, excluding non-wanted characters, and, by consequence, the duplicated articles. Other problem was collecting the *dates* of the articles. Using the "Selector Gadget" programme was not enough because the dates of publication, in many of the pages, were related to more than one article and, when scrapping, it just showed for the main article published in that day. Hence, I used the URLs to pull the dates, by an extraction function. In this first part I scraped Titles, Dates and URLs.

Here, I am not interested in all articles about Germany, but only the ones where AfD is referred. To create a data frame with only articles about AfD I filtered using a *grepl* function. I pulled all articles containing related topics to the party: “AfD”, “German far-right party”, “Alternative für Deutschland” and “Alternative for Germany”. I did not use the term “far-right” because this could return non-wanted results, like news about other far-right groups and parties. Using this filter, I obtained 45 articles.

The third part was to pull the full articles, the main data in this research. I pulled this content by reading the html of my URLs. Later I cleaned the raw data because it came with non-wanted objects, such as references to other news, “Read more” and labels to share on social medias. 4 out of 45 articles were videos or images. So I deleted them from my data frame, avoiding problems by not having texts in all the articles. In the end, The Guardian had 41 articles about AfD.

b. The Data Collection in Deutsche Welle newspaper: scope and challenges

The collection of DW articles was also automated. However, it was a more challenging task to do. The international website of the newspaper does not contain a label about Germany where I could go from page to page to extract the articles. The label about Germany contain only the “top stories”. Hence, I typed “Germany” in the search and select the time frame. But, the page does not go from page 1 to 2 and to others, so I couldn’t loop as in The Guardian. The page has a “show next 10 results” label and it works only until the search shows the first 1000 articles, even if it is asked for more.

Thus, I looped considering the dates. I used the *while* function to change the URL for each 1000 articles, considering my starting and final date. I used “Selector Gadget” to get the nodes for the Titles, Dates and URLs. Even with this process I had some duplicated articles and the titles came with the dates attached. So, I used the *stringr* package to clean the titles and the *unique* function to create a data frame without duplicates. My last problem in this phase were the URLs. It was coded differently. The URLs didn’t come complete so I used a paste0 function to add the “http://www.dw.com”. I got 32,667 results in a data frame.

Not all these articles were interesting, since I used a broad search word. Thus, I used the same labels as in The Guardian to create a data frame for articles about AfD, resulting in 363 articles.

With the URLs organized I pulled all the content, getting the entire articles. After, I used a *gsub* function to remove all the non-wanted characters or phrases, such as “Send on Facebook”. The newspaper also contained some news only with images or videos and I used a *grepl* function to remove them. I ended this phase with 344 articles with full text.

3 Data Analysis

The procedures for data analysis were the same for the two articles. The purpose of this paper is to observe how the newspapers portrayed the party through the time analyzed. Further, I aim at looking if the phases of transition and shifts in the agenda reflected in the articles. In this sense, my research is focused on the 4 phases observed by the study of the historical background of the party. I selected three analysis to be done with the data collected: topic models, K-means and Naïve Bayes.

a. Topic Models

I used the *quanteda* and *tm* package to clean the articles. I checked the top features and the mostly typed words in all the articles. Some words had to be removed from my analysis to guarantee an observation of my data without noise. To illustrate, the image shows the top features for a non-cleaned data from DW.

afd	parti	germani	german	said	right	state	elect	merkel	year
3823	3170	2092	1502	1364	1214	1002	944	774	752

This isn't interesting, because it is trivial that AfD, for instance, would be one of the most quoted words, since all the articles are about this theme. I created a set of words to be excluded from the data, i.e. AfD, Germany, names of social media. For The Guardian newspaper I observed the same words from the graph, only in different amounts.

After the cleaning, I ran the topic models code. To check the frequency of the words, I used the *term frequency-inverse document frequency* (tf-idf) to show the importance of each word. Thus, I checked the importance of the term in the corpus of the document, instead of only its frequency. This is because checking a frequent term by itself does not necessarily mean its importance (Imai, 2017).

For the topic models I programmed to divide the articles into four topics, according to what the programme found as the most relevant after using the *tf-idf* function. These clusters were random and demonstrated which articles fitted the best in each cluster.

After randomly sample the data, 250 articles were used to train and test the articles' clusters. The programme checked the optimal number of clusters, generating other way to group by topic. I run the top features function for the phases I created. My goal with this procedure was to compare the top words with the expected theme in each cluster.

Using topic models, I checked what were the top features in each of the clusters defined by AfD's trajectory. By running a topic model in the articles, I checked if there were other possibilities of clustering and how different or similar they were in relation to the one I had. Lastly, checking for the optimal number of clusters and its topics helped to check other ways of identifying the related themes.

b. K-means

For the K-means analysis I *quanteda* and *tm* to clean the text. The k-means analysis is the classification of certain object according to the characteristics of its neighbors. In other words, the nearest neighbors' classifiers without a label are defined by their characteristic, have a class assigned to them according to the similar labeled examples. It is necessary to divide the data into train and a test categories, randomly. Thus, the algorithm is trained with the labeled articles and applied in the to test part (Lantz, 2013).

The goal is to observe how, through the package *clusters,* the articles are classified. Hence, I create a Document Feature Matrix with the top features after applying the *tf-idf*. I made the clusters using the k-means formula and checked if it fits.

The goal here was to observe how the programme would classify the articles into the clusters. I wanted to see if what I learned from the studies on the historical background of the party would meet the way these two newspapers reported about the party.

c. Naïve Bayes

The last analysis mechanism was a probabilistic learning – the classification using Naïve Bayes. This is a machine learning algorithm using principles of probability for classification. Hence, it uses the data of prior events or classifications to estimate the probability of future events (Lantz, 2013). In this article, the algorithm is a way to classify the articles based on the prior knowledge of classification acquired by the study of the historical background of the party.

I used the package *tm* to clean my data and the package *e1071* for the analysis. First I randomized my dataset. This is because the Naïve Bayes procedure is done by splitting my data into train and test. For that I used the rows and split them in 75% for training and 25% testing.

The rows as in the original data frame did not work because they are selected through a time line and training and testing would be biased by time.

Train and test were also shaped in a DTM format. The function *dictionary* controlled the data with only the words that appeared in the frequent terms expression, at least 5 times.

The algorithm together with a function classified the articles. One of the difficulties for this part was to make my data work. I had some barriers when doing the function because of the format of my dataset. Thus, I was able to do by transforming my data into a data frame and into factors.

Hence, using the Naïve Bayes algorithm I could predict the classification of the articles using a training model. With this and applying to my entire data set, I observed if the algorithm can understand the way the articles were clustered and if it follows the same path with the testing.

4 Results

The articles' analyses were made separately from one newspaper to other in order to see if there are differences between the way the party is portrayed in the domestic news and at the international level. All results were obtained with coding procedures in R programme. I aimed at answering these questions with the analyses:

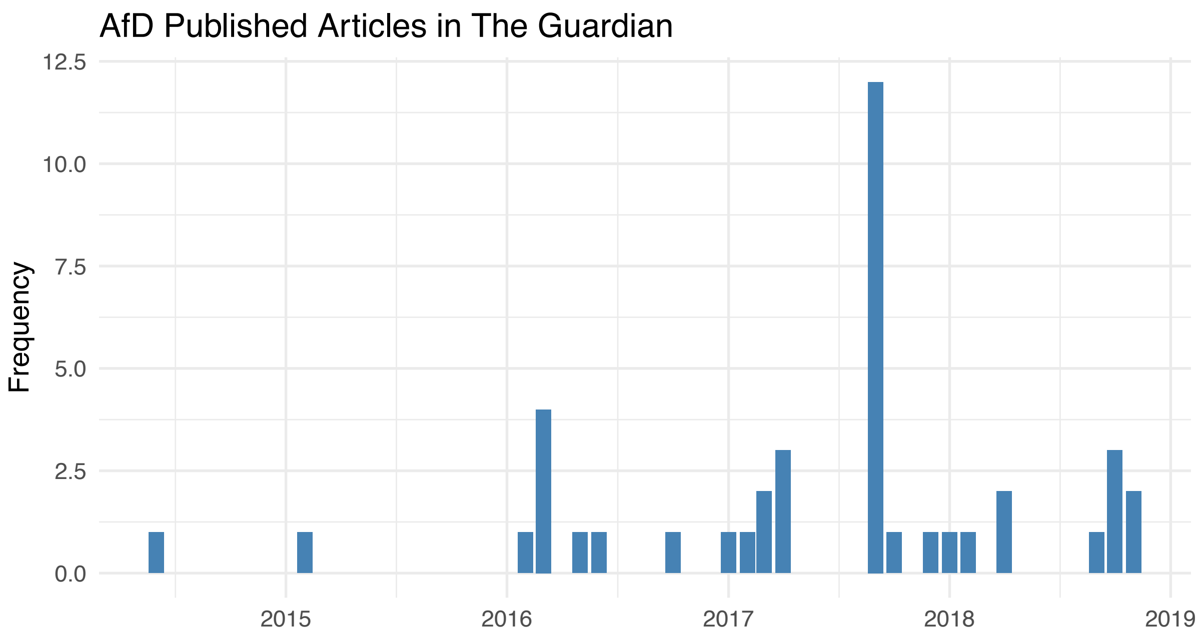
Q1: What is the overall portray of the party to the newspapers' readers? What do they most read about AfD?

Q2: Regarding to the topics, is a time clustering an efficient way to demonstrate the shifts in the agenda of the party through the news? Is there any other way of cluster these articles?

a. The results from The Guardian

I analyzed 41 articles. It is interesting to observe that the first article published was only in June/2014. It shows that the first appearance of AfD in The Guardian newspaper was after the European Parliament elections, where the party was already credited with some seats and gained relevance. Between 2014 and 2017 there were not that many publications about AfD.

However, in September 2017 the publications about AfD increased considerably. More than ¼ of the data was published in that period.

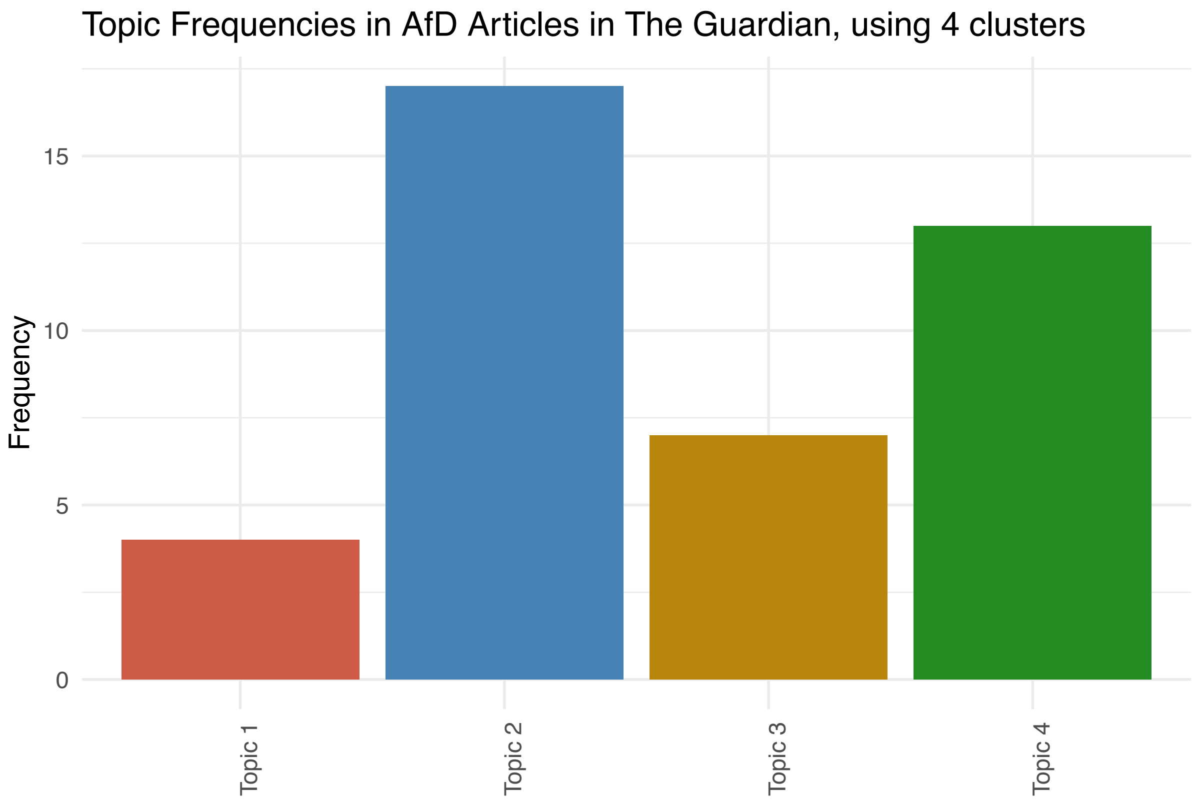


In September 2017 The Guardian published 12 articles about the party. In the following months, until November 2018, the newspaper published 24 articles about the party. This means that in less than one year they mentioned AfD in more than 50% of the total mentions about the party in almost six years. There are two explanations: the party was important for the Bundestag elections in 2017 and now is the third largest party and the main opposition. Therefore, it is understandable this raise in the publications.

i. Topic Models

The articles were clustered into four topics.

	Topic 1	Topic 2	Topic 3	Topic 4
[1,]	"vote"	"state"	"peopl"	"right"
[2,]	"right"	"elect"	"labour"	"elect"
[3,]	"elect"	"cdu"	"european"	"weidel"
[4,]	"east"	"vote"	"conserv"	"peopl"
[5,]	"poll"	"refuge"	"member"	"far"
[6,]	"far"	"support"	"polit"	"world"
[7,]	"state"	"coalit"	"social"	"war"
[8,]	"voter"	"result"	"right"	"member"
[9,]	"saxoni"	"democrat"	"left"	"höcke"
[10,]	"oppach"	"polit"	"mep"	"parliament"
[11,]	"open"	"countri"	"gauland"	"bundestag"
[12,]	"chancellor"	"parliament"	"econom"	"email"
[13,]	"refuge"	"poll"	"storch"	"gauland"
[14,]	"part"	"anti"	"ecr"	"polit"
[15,]	"peopl"	"spd"	"bank"	"protest"
[16,]	"democrat"	"polici"	"centr"	"farag"
[17,]	"support"	"voter"	"hitler"	"confer"
[18,]	"immigr"	"populist"	"far"	"populist"
[19,]	"lead"	"bundestag"	"berlin"	"berlin"
[20,]	"won"	"govern"	"nazi"	"septemb"



The Topic 4 contains words like “elect”, “bundestag”, “populist”, “höcke” and “septemb”. These are words referring to the elections and to people related to the party. In this sense, we can observe that the Topic 4 is about elections and the presence of the party in the

German Politics, more specifically the last elections. This was expected, because the importance of the party changed significantly from 2017, so it is understandable that there is a topic where political words are the majority. It is interesting to look in the Topics 2 and 3. They contain words related to the migration crisis and to the euro crisis. This is illustrated by “European”, “immigr”, “refuge”, “econom”, “bank”. The topic 1 is also related to politics and voting procedures, as there are words as “bundestag”, “anti”, “poll” and contains some traces about immigration matters.

In the Graph we observe the topic 2 with the highest amount of articles. The topic contains words related to the immigration crisis and to elections. By analyzing the history of the party we observe that after the shift to the immigration crisis the party gained more importance in the political scenario. This was also the period when it became the third largest party. Thus, it makes sense that the majority of the articles is about these topics.

11 was the optimal number of clusters found in the analysis, when I randomly sampled the data set. It is noticeable that some of the topics referred to the European Parliament elections and to eurozone issues. Other topic was related to “UKIP” and to “bundestag”, probably mentioning the presence of far-right parties in Europe. Words referring to refuge and immigration were also present.

To answer Q1, the Topic Models showed that the high majority of news are exactly about what the party has been dealing through the years. The most recurring themes are related to elections and to the two EU crises related. “muslim” and “refuge” terms appear more often, maybe because the party became more relevant after adopting immigration related agendas.

Regarding to Q2, it seemed efficient to make four clusters. A Cross table do not show the clusters labeled as the phases. However, when I analyze how the programme grouped the articles, it kept almost the same articles into the same clusters. For the topic models, using the optimal k also bring interesting results but the topics overlap regarding the themes.

I also analyzed the Top Features phase by phase, using the division I made according to the literature. There were none articles published in the Phase 1, probably because the party was not significant enough for the international scenario. For the Phase 2, the most typed words were related to elections, bank, and seats, relating to monetary and electoral issues. Phase 3 contain words about refuge and politics: this was expected. Phase 4 has a a dominance of words related to party and elections, probably due to the 2017 Elections. The results matched with the predictions.

ii. K-means

Through k-means the result was a division of 7 articles in the Cluster 1, 3 in the Cluster 2, 1 in the Cluster 3 and 30 in the Cluster 4. When observing the centers, “polit”, “right”, “rightw”, “European”, “public” and “conserve” had the highest values, so these terms guided the clustering process. However, K-means only classified the articles regarding the political reasons, as the elections. It did not recognize more specific clusters, as the ones related to the crises.

For this analysis, the overall portray of the party is related to elections and politics. The model could not recognize more specific topics and, therefore, the answer for the Question 1 is a scope based in the political procedures, without taking into account the shifts in the party agenda or its development through the years.

For the Question 2, the clustering suggested for the algorithm refers to aggregate all the news about politics in one cluster only. However, this seems to minimize the agenda topics of the party and focus in the party itself, as an agent in the German political scenario.

Hence, k-means worked but did not give any relevant answers for the research questions.

iii. Naïve Bayes

For Naïve Bayes there were 30 articles for training and 11 articles for testing. The articles for training were a classifier, according to the Phases developed previously. The programme predicted the phases for the 11 testing articles, according to the classifier. In the testing part there was no articles for Phase 2. This is because in this newspaper the Phase 2 had only 2 articles. So, when the programme predicted it did not used the Phase 2.

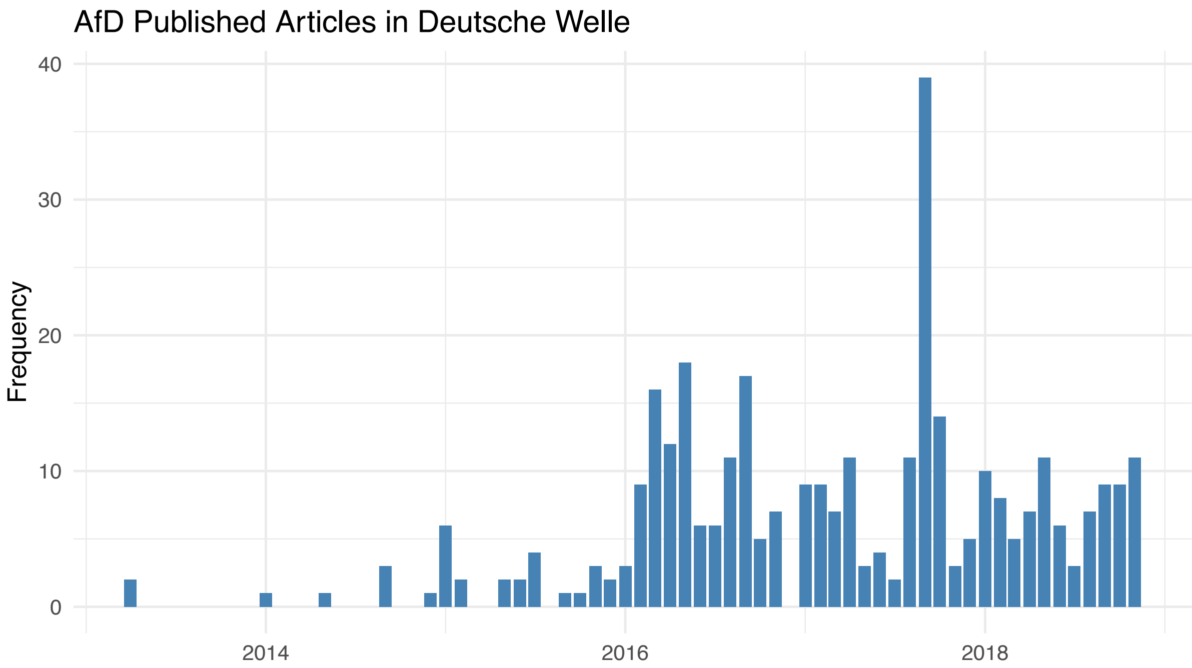
The programme predicted 8 articles in the Phase 3, in which 7 of them matched with the previous classification. Out of the 11 articles, 3 were predicted in the Phase 4, while 2 of them were already classified there.

```
   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  41


             | afd$Predictions
  afd$Phases |         2 |         3 |         4 | Row Total |
-------------|-----------|-----------|-----------|-----------|
           2 |         0 |         1 |         1 |         2 |
             |     0.293 |     0.005 |     0.211 |           |
             |     0.000 |     0.500 |     0.500 |     0.049 |
             |     0.000 |     0.045 |     0.077 |           |
             |     0.000 |     0.024 |     0.024 |           |
-------------|-----------|-----------|-----------|-----------|
           3 |         4 |        17 |         6 |        27 |
             |     0.001 |     0.436 |     0.766 |           |
             |     0.148 |     0.630 |     0.222 |     0.659 |
             |     0.667 |     0.773 |     0.462 |           |
             |     0.098 |     0.415 |     0.146 |           |
-------------|-----------|-----------|-----------|-----------|
           4 |         2 |         4 |         6 |        12 |
             |     0.034 |     0.924 |     1.266 |           |
             |     0.167 |     0.333 |     0.500 |     0.293 |
             |     0.333 |     0.182 |     0.462 |           |
             |     0.049 |     0.098 |     0.146 |           |
-------------|-----------|-----------|-----------|-----------|
Column Total |         6 |        22 |        13 |        41 |
             |     0.146 |     0.537 |     0.317 |           |
-------------|-----------|-----------|-----------|-----------|
```

The Naïve Bayes prediction system was one of the most interesting for answering the research question. As observed in the Cross Table above, the programme predicted the articles very similar to the phases. Hence, the Naïve Bayes algorithm answer us the Question 2. It shows that through prediction, by the training of part of the previous classification, we can estimate a similar classification to the one made before. In other words, the algorithm worked to show a time cluster in a very similar way as the time line. This algorithm does not tell us anything about how it portray because it do not analyze topics, by showing them. It only shows the predictions of the clusters.

b. The results obtained from Deutsche Welle

The Phase 1, until 31.05.2014, had only 4 articles. The Phases 3 and 4 represent 90% of the total amount of articles.

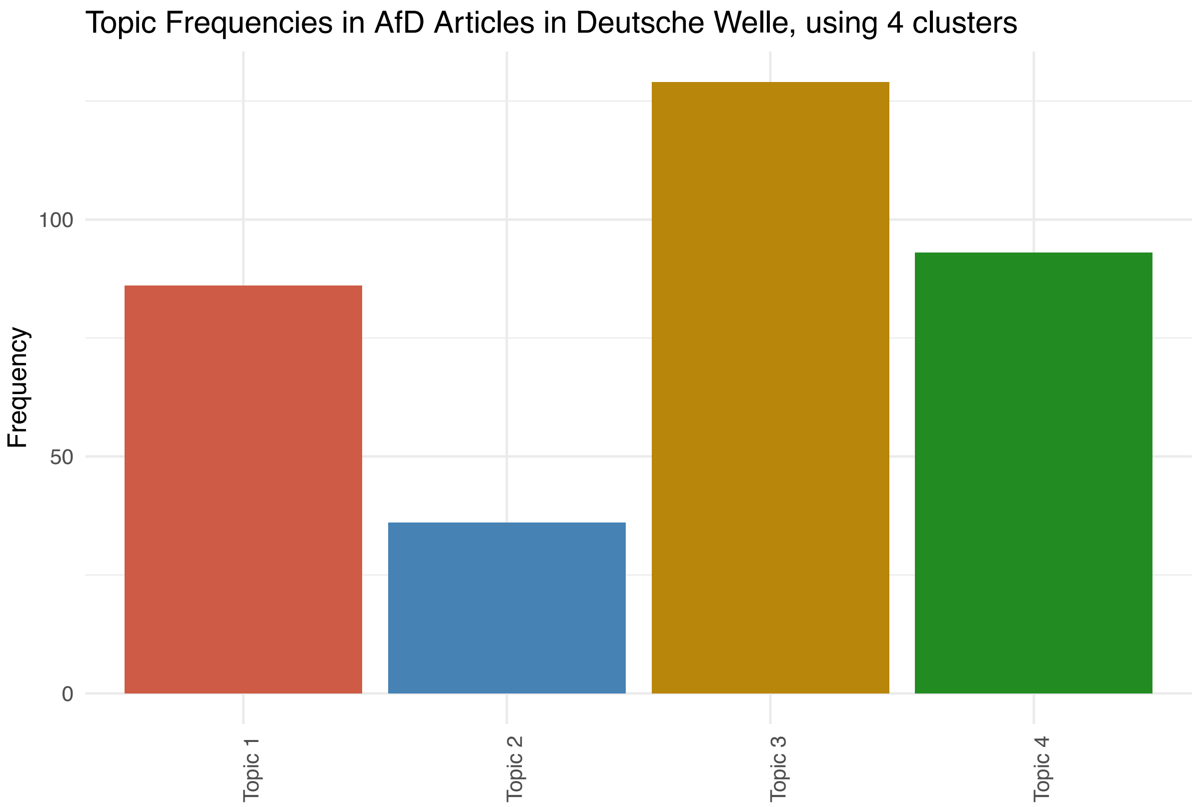


This considerable raise in the publications number in the last two phases follows the same reasons for The Guardian: the importance of the party for the 2017 Elections and the gains of visibility of the party around this stage.

i. Topic Models

The articles were clustered into four topics.

```
      Topic 1      Topic 2      Topic 3      Topic 4
 [1,] "right"      "gauland"    "petri"      "cdu"
 [2,] "bundestag"  "state"      "anti"       "elect"
 [3,] "peopl"      "weidel"     "right"      "refuge"
 [4,] "far"        "nazi"       "state"      "spd"
 [5,] "polit"      "border"     "member"     "democrat"
 [6,] "old"        "refuge"     "elect"      "state"
 [7,] "member"     "co"         "nation"     "voter"
 [8,] "read"       "right"      "polit"      "right"
 [9,] "wing"       "wing"       "islam"      "vote"
[10,] "nazi"       "elect"      "wing"       "social"
[11,] "can"        "attack"     "höcke"      "left"
[12,] "time"       "made"       "frauk"      "support"
[13,] "left"       "politician" "immigr"     "chancellor"
[14,] "parliament" "far"        "parliament" "coalit"
[15,] "cultur"     "member"     "populist"   "poll"
[16,] "use"        "boateng"    "luck"       "green"
[17,] "elect"      "minor"      "european"   "polit"
[18,] "media"      "illeg"      "europ"      "polici"
[19,] "critic"     "alexand"    "eu"         "govern"
[20,] "protest"    "remark"     "support"    "csu"
```



The Topic 1 contains words like "bundestag", "far", "right" and "elect", referring to elections and political disputes, which is expected. Topic 2 contains "refuge" and "border", referencing to the immigration crisis, an important topic for AfD's agenda. The topic 4 is a reflection about the positioning of AfD in politics, by the words: "voter", "right" and "poll". Checking the graph for the top frequencies gives us an interesting observation: the Topic 3 had the highest amount of articles. This topic is also a mix of all the party agendas and about elections. In this sense, the programme made the Topic 3 as a generic topic where there is a mixture of all the agendas related to the party.

2 was found as the optimal k. It followed the same procedure as The Guardian. For the Topic 1, "vote", "democrat" and "parliament" appeared, referencing to politics and elections. Topic 2 was related to the refuge crisis and nativist sentiments, with "border", "nazi", "extrem" and "refuge". The division is comprehensible because for almost 3 years the party adopted such agenda.

For Q1, the Topic Models showed that the features in the topics were also the themes involving AfD through these years. It was related to the two euro-crises, also in their agenda, as well as elections. Regarding to Q2, it seemed efficient to make four cluster for Deutsche Welle. The way with two topics also reflects the trajectory of the party and its agenda, even though it masks the original agenda of the AfD.

For the Phase 1the top features were related to "euro", "European", "member" and "state". This portray partly the agenda for the party in that time, where they were against the Eurozone. Phase 2 contains "pegida", "right", "luck" and "European", reflecting the change in the leadership of the party, which happened in the same time. Phase 3 is a mix between voting procedures and the agenda on immigration crisis, by the words "petri", "refuge", "anti", "vote". Phase 4 is related to voting process with "far", "elect", "polit", "bundestag" and "wing". In this sense, it makes sense the clustering division with the background of the party. The top words matches with what was expected.

ii. K-means

32 articles in the Cluster 1, 237 in the Cluster 2, 3 in the Cluster 3 and 72 in the Cluster 4 represented the k-means clustering. When observing the centers of the k-means, "right", "wing", "youth", "elect", "refuge" and "asylum" had the highest values, the centers for clustering. Differently from The Guardian, in Deutsche Welle the words related to the agenda of the party are influencing the clustering. The k-means recognized more than only electoral issues, but that was still minimum.

Regarding the k-means modelling, the overall portray of the party is related to elections and politics, with traces of the political agenda. Therefore, the readers mostly read news about elections with traces of immigration issues. This difference comparing to The Guardian is understandable because DW is a domestic newspaper, so the articles about local news are more specific.

For the Q2, the clustering suggested reference to aggregate partly the news with electoral issues. It is interesting that the cluster 1 aggregated news about radicalization, such as the article n. 150 about pepper spray or the n. 35 about radicalization itself.

The k-means was more effective in Deutsche Welle comparing to The Guardian results but still did not achieved a proper answer for the research questions.

iii. Naïve Bayes

258 articles were used for training and 86 for testing. The procedure was the same as in The Guardian. The results for testing:

	Classified Phases		
Predicted Phases	2	3	4
1	0	1	0
2	0	0	1
3	4	44	18
4	0	5	13

For the testing phase, the programme classified 44 articles in the Phase 3 as I did in the beginning of the article, out of 66, meaning 66% of a right prediction. For the Phase 4 it was 72% of right prediction.

```
   Cell Contents
|-------------------------|
|                       N |
| Chi-square contribution |
|           N / Row Total |
|           N / Col Total |
|         N / Table Total |
|-------------------------|


Total Observations in Table:  344


             | afd$Predictions
  afd$Phases |         1 |         2 |         3 |         4 | Row Total |
-------------|-----------|-----------|-----------|-----------|-----------|
           1 |         0 |         1 |         3 |         0 |         4 |
             |     0.140 |     0.000 |     0.053 |     0.244 |           |
             |     0.000 |     0.250 |     0.750 |     0.000 |     0.012 |
             |     0.000 |     0.012 |     0.013 |     0.000 |           |
             |     0.000 |     0.003 |     0.009 |     0.000 |           |
-------------|-----------|-----------|-----------|-----------|-----------|
           2 |         1 |         4 |        13 |         2 |        20 |
             |     0.131 |     0.180 |     0.001 |     0.497 |           |
             |     0.050 |     0.200 |     0.650 |     0.100 |     0.058 |
             |     0.083 |     0.047 |     0.058 |     0.095 |           |
             |     0.003 |     0.012 |     0.038 |     0.006 |           |
-------------|-----------|-----------|-----------|-----------|-----------|
           3 |         6 |        51 |       142 |        13 |       212 |
             |     0.263 |     0.037 |     0.053 |     0.000 |           |
             |     0.028 |     0.241 |     0.670 |     0.061 |     0.616 |
             |     0.500 |     0.600 |     0.628 |     0.619 |           |
             |     0.017 |     0.148 |     0.413 |     0.038 |           |
-------------|-----------|-----------|-----------|-----------|-----------|
           4 |         5 |        29 |        68 |         6 |       108 |
             |     0.403 |     0.201 |     0.123 |     0.053 |           |
             |     0.046 |     0.269 |     0.630 |     0.056 |     0.314 |
             |     0.417 |     0.341 |     0.301 |     0.286 |           |
             |     0.015 |     0.084 |     0.198 |     0.017 |           |
-------------|-----------|-----------|-----------|-----------|-----------|
Column Total |        12 |        85 |       226 |        21 |       344 |
             |     0.035 |     0.247 |     0.657 |     0.061 |           |
-------------|-----------|-----------|-----------|-----------|-----------|
```

The system was interesting. The testing part achieved results similar to the ones previously classified, accompanying the logic behind the phases. Using all the data, including the testing data, most of the predictions were correct, as it is possible to see in the Cross table above. We can observe a focus on the phase 3, that is the one with most articles. In the first classification there were 212 articles and in the prediction model the programme classified 226 articles. It was a good average result for the analysis.

The clusters in the algorithm were similar to the ones from the timeline. Therefore, the algorithm is able to answer the Q2, demonstrating, through training and testing, similar results.

There are no answers for Q1 regarding topics, because it only shows the predictions based on my model. So the Naïve Bayes only works to confirm my model and to suggest different allocations for certain articles.

Conclusions[1]

The models chosen for the analysis helped to answer the research questions. Topic Models analysis was certainly the best option for understanding how the themes in the articles related to the shifts in the party agenda. It was interesting to check the top features of each phase because it proved that the historical background affected the portray of the party in both the international and domestic media.

K-means, by the other hand, was the most inefficient method considering the research questions. It worked better in Deutsche Welle comparing to The Guardian and this may be explained by the amount of articles and the domestic focus of the first newspaper. However, it presented another way of clustering articles that also complied with the profile of AfD.

Naïve Bayes algorithm proved to be interesting in order to check about the predictions of clustering made during the study of the history of the party. It did not present me an analysis of the topics, but helped to substantiate the choice for the four phases.

The *Alternative für Deutschland* is a phenomenon for political science. The party gained considerable grounds in just a few years of existence and has been influencing other parties, with the same bias, in other European countries. This is the reason why it is relevant to research about it. The news also played an important role for the increase of the AfD's electorate. Hence, this article contributed to understand the way the party was portrayed in the media through these years. It helped to understand if the changes in the agenda were reported in the articles.

[1] The LaTex for producing pdf versions in R did not work in my computer, so I opted by writing the article in word format and then formatting to pdf

References

Art, D., 2013. Rise of the Radical Right: Implications for European Politics. Brown J World Aff 19, 127.

Arzheimer, K., 2015. The AfD: Finally a Successful Right-Wing Populist Eurosceptic Party for Germany? West Eur. Polit. 38, 535–556. https://doi.org/10.1080/01402382.2015.1004230

Berning, C.C., 2017. Alternative für Deutschland (AfD)-Germany's New Radical Right-wing Populist Party. DICE Rep. 15, 16–19.

Chazan, G., 2017. German media accused of one-sided coverage of refugee crisis [WWW Document]. Financ. Times. URL https://www.ft.com/content/23e02b76-7074-11e7-93ff-99f383b09ff9 (accessed 1.4.19).

Goerres, A., Spies, D.C., Kumlin, S., 2018. The Electoral Supporter Base of the Alternative for Germany. Swiss Polit. Sci. Rev. https://doi.org/10.1111/spsr.12306

Grimm, R., 2017. The rise of the German Eurosceptic party Alternative für Deutschland, between ordoliberal critique and popular anxiety. Int. Polit. Sci. Rev. 36, 264–278.

Imai, K., 2017. Quantitative Social Science: An Introduction. Princeton University Press, Princeton & Oxford.

Knight, B., 2018. Is Germany's media to blame for the inexorable rise of the AfD? | Ben Knight. The Guardian.

Lantz, B., 2013. Machine Learning with R. Packt Publishing, Birmingham.

Lees, C., 2018. The 'Alternative for Germany': The rise of right-wing populism at the heart of Europe. Polit. Polit. Stud. Assoc. 38, 295–310. https://doi.org/10.1177/0263395718777718

Schmitt-Beck, R., 2017. The 'Alternative für Deutschland in the electorate': between single-issue and right-wing populist party. Ger. Polit. 26, 124–148. https://doi.org/10.1080/09644008.2016.1184650