# Interdisciplinary Project in Data Science

Analysis of the Twitter Conversation Under the Hashtag #IchBinHanna

Ahmadou Wagne
TU Wien
Vienna, Austria
12002293

Main Supervisor:
Dr. Jana Lasser
TU Graz
Graz, Austria

Co-Supervisor:
Prof. Allan Hanbury
TU Wien
Vienna, Austria

## ABSTRACT

The general objective is to perform sentiment analysis and topic modelling on a collection of tweets under the hashtag #IchBinHanna. The conversation under this hashtag is about the precarious working conditions of researchers especially in Germany. Dictionary based sentiment analysis approaches VADER and LIWC are used to observe a rather positive overall sentiment in the time of June to July 2021. For topic modelling LDA is used to see which topics are prominent at which point in time. Finally different user groups related to the academic system are included into the analysis and their sentiment, as well as the topics they tweet about are compared. This project introduces preparation steps like including hashtags and mentions into the topic models, that lead to valuable results. It also reveals some of the shortcomings of popular methods, like the misclassification of ironic tweets, which are prominent in this specific discussion, by dictionary based sentiment analysis methods.

## KEYWORDS

sentiment analysis, natural language processing, topic modelling, micro blogging, twitter, #IchBinHanna

## 1 INTRODUCTION

Hanna is a fictional character that originated from a video[1] of the German Federal Ministry of Education and Research. The video should describe the 'Wissenschaftszeitvertragsgesetz'[2](further referenced as WissZeitVG), which is a German law that allows institutions employing researchers to give temporary contracts to scientific workers. This is furthermore a problem, because the total duration for being employed under such contracts is limited (to nine years for post doctoral researchers), which means that there is no chance to stay in the system, if one has not managed to get a tenured position during this short period of time. Hanna is an example for a scientific worker employed under such a contract, while writing her doctoral thesis. The video tells that this law exists so that people like her, who are in the system for a longer time, do not block the jobs and resources for younger generations. The clip also indicates that although the contracts are temporary, they have to be long enough to get the desired degree. This caused a great outrage among a lot of scientists especially in the German speaking DACH region. As a spontaneous reaction Amrei Bahr, Kristin Eichhorn and Sebastian Kubon created the hashtag #IchBinHanna, which references the character of said video. The first relevant tweet of Sebastian Kubon that started the trend can be seen in Figure 1.


**Figure 1: First #IchBinHanna tweet by Sebastian Kubon [3][13]**

Under this hashtag, scientists of different backgrounds began to tell their stories and made it clear that for them, the WissZeitVG is not beneficial at all and that they feel taunted by the display of Hanna's situation with whom they can identify. The law causes a great amount of problems for scientists and their lives . Currently around 92% of scientific workers at universities are employed on a temporary contract. Some claim that it is impossible to plan one's life or career, under the constant pressure to get a temporary contract somewhere in Germany to be able to stay in the system. It is also not rare that people in their mid-thirties already signed five or more temporary contracts [12]. In addition to get a tenured position a high number of publications is considered a necessary condition. As a consequence, chances are there that the quality of scientific work suffers, as workers might be pushed to publish as much as possible to have a chance at continuing their academic career.
All these stories told on Twitter led to increased attention for the topic. As a first reaction, the ministry deleted the video from their website. Since the emergence of the hashtag around the 10th of June 2021, there was also a lot of media coverage[18][32][29] and responses to the movement.
This project aims at giving insight on how the conversation on Twitter unfolded in the time between June 2021, when the hashtag first occurred, and the end of September 2021 after the federal elections in Germany. Over this period, a collection of all tweets posted

[1]https://web.archive.org/web/20210611145015/https://www.bmbf.de/de/media-video-16944.html
[2]https://www.gesetze-im-internet.de/wisszeitvg/BJNR050610007.html
[3]Consent to print the tweet given by Sebastian Kubon

under the hashtag #IchBinHanna is analyzed. In the following, the project takes two main perspectives on the conversation on Twitter, brings them together and shows the development over time. On the one hand sentiment analysis is performed to try to catch the mood expressed in tweets by people participating in the conversation on the topic. As the discussion is dominated by people describing their precarious working conditions, tweets are intuitively assumed to express a rather negative sentiment overall. The first question raised here is, whether the methods applied can validate that negative tweets dominate in the conversation. With increasing attention and given the complexity of the issue and potential solutions, the conversation most likely evolved significantly over time. To understand what people actually talk about at a given point in time, the second perspective is about identifying events and topics related to the hashtag. This is done by a combination of qualitative and quantitative data analysis approaches (e.g. observing spikes in tweet volume at a certain day) and topic modelling.

As the aim of the movement is to raise awareness and eventually cause a change in the way the academic system works, it is interesting to track the development of the conversation over time. A goal is also to tie the derived sentiment to the identified events, as for example a news report or the opinion of a politician can trigger different reactions from affected persons and they will most likely express this in the way they talk about the event on Twitter. Also the sentiment can be an indicator in whether the movement is content with the impact it made. In parallel to this project Ellen Le Foll[4] developed a linguistic model to detect user profiles of different types of Twitter users that are in some way affiliated to academics or are relevant in the discussion of the issue (like media or politicians). The user classifications are used in the analysis presented here to be able to compare sentiment and the topics discussed between groups and also to get a picture of who took part in the conversation.

This leads to the following research questions that are central to this project:

- Is the overall sentiment of the #IchBinHanna-movement on Twitter negative?
- Which are the most important topics and events that can be detected in the tweets discussing #IchBinHanna?
- What is the observed sentiment of the conversation during the time detected events are discussed?
- Does sentiment and the topics discussed vary between groups?

The following analysis was conducted as part of the 'Interdisciplinary Project in Data Science' of the Data Science curriculum at TU Vienna and the code is available in a public Github repository[5]. The exact underlying data is not publicly available, but the tweet IDs were published to reproduce the results[6].

# 2 RESEARCH METHODS AND RELATED WORK

The work is subdivided into three parts. The first part is a descriptive data analysis, that focuses on raw statistics of the relevant data set. It uses tweet volumes and word frequencies to identify topics discussed at a specific time with an exceptionally high output of tweets containing the hashtag. The exact steps that are performed are described in the section 3.1. The second part is sentiment analysis, described in section 2.1 and the last part in section 2.2 is topic modelling using latent Dirichlet allocation to identify further topics and analyse the sentiment tied to those topics in section 4.4.

## 2.1 Sentiment Analysis

Sentiment analysis is an efficient and widespread method to extract subjective information from raw text. The goal is to detect the sentiment of opinions/texts and classify it for example in a binary case into positive or negative sentiment. At the start of the 2000s, after the rise of the world wide web, it was originally used to process web reviews[19], web blogs[16] or news articles [10] to gather public opinion on a larger quantified scale, as more and more data was available. Since then methods have advanced a lot, gained a lot of interest and branched out into different fields. Soon approaches were introduced to identify a more fine grained selection of classes of affective states, like emotions (sadness, anger etc.)[27] and have developed to more sophisticated approaches with the application of neural networks [2]. With the breakthrough of social media, a vast amount of textual data generated by individual people was produced and a lot of it is publicly available. In particular sentiment analysis methods using machine learning now got access to a lot of training data.

This project deals with micro blogging data from Twitter specifically, which comes with interesting features, but also a lot of difficulties in processing it. Tweets serve to express oneself and are often used as an online representation of a person's beliefs and opinions. Moreover if many people share their opinion on the same topic, it is possible to follow occurring trends almost in real time. Especially because tweets are enriched with additional semantic information liked hashtags, it can be easy, but not trivial, to extract tweets concerning a certain topic. In that way, the data set used here was retrieved by using the Twitter API v2 to retrieve tweets containing the hashtag #IchBinHanna.

Tweets contain textual information with up to 280 characters (140 until 2017) and can additionally include media like images or videos. They can also be written as a reply to another tweet or 'cite' a tweet as a retweet (with or without an added comment). There is only a short amount of characters to express one's opinion, which makes it a challenge for all kinds of natural language processing tasks. What is also challenging is that the writing style on Twitter is very different to news articles for example, which are typically reviewed by another journalist. Tweets often contain specific language like slang or abbreviations and emojis. Another problem can be the elongation of strings like 'nooo' instead of 'no' or simply spelling errors. Some of those aspects can hold information that can be used for sentiment analysis like emojis for example, but there are many different kinds of methods and opinions on how to use them [7]. A major challenge also faced in this project is the detection of sarcasm or irony, which makes it especially difficult on a word level, as positive words would in that case sometimes lead to a negative sentiment and vice versa. The last challenge that lead to the decision of the approaches used for this project is that methods available for sentiment analysis heavily dependent on the language

[4]https://github.com/LaserSteff/IchbinHanna/blob/main/code/UserClassification.Rmd
[5]https://github.com/ahmadouw/IchBinHanna-Analysis
[6]https://bit.ly/3o0mbYN

of the underlying data.
In general there are two main approaches to conduct sentiment analysis: machine learning, which mostly happens in a supervised setting, and dictionary based approaches. Supervised machine learning models achieving good performance use classical algorithms like Naive Bayes [26] or Support Vector Machines [21], but also developed in the direction of neural networks and transformer based pre-trained models [1]. For those approaches there is still a lack of good German training data that specifically fits a given task. Furthermore, subjective texts are domain dependent for different topics (even within the same platform) and manually annotating data is very resource consuming. As the available data is of course not labelled and no work concerning sentiment analysis has been done for the topic of #IchBinHanna, the decision was made to use the second type of approach, namely dictionary or lexicon based methods. They also make it easier to interpret results later, as they are fully comprehensible by looking at the underlying word lists used to generate scores. Those methods use a dictionary containing words that are determined relevant towards a specific kind of sentiment (opinion words). They have to be collected and annotated beforehand, but a lot powerful resources like WordNet, which is used to develop a Thesaurus called SentiWordNet [9], are already available, showed great performance and are regularly updated. As most of the tweets here are German, the decision on available dictionary based methods fell on Linguistic Inquiry and Word Count (referenced as LIWC) [20], as well as Valence Aware Dictionary for Sentiment Reasoning (referenced as VADER) [11]. As there is no data available to validate the results, two approaches are selected to see how they agree on their classifications.

#### *2.1.1 LIWC.*

LIWC is a text analysis software that was initially developed in 2001 for the field of psychology and was psychometrically validated by creator Jamie Pennebaker among others. Since then it has also been applied in various fields like sentiment analysis for micro blogging posts[24], marketing[15] or mental health evaluation[31] and the support for many languages was added, by adding lexicons for German, Arabic, French and many other languages. There are various releases of the dictionaries as they are updated. They contain a list of words that are assigned to one or more classes like 'function words', 'affect', 'informal language' etc. These labels can be used to profile or identify writing styles, mark some part of speech tags, identify specific topics or classify texts as positive or negative. As the only tweets that are relevant for this analysis are either German or English, the German LIWC 2007 and the English LIWC 2015 are selected and loaded via the Python implementation of the LIWC software[7]. The only manual modification made was that the word 'like' was excluded from the English version, as it is by experience not very useful when using this dictionary based solution, due to its semantic ambiguity. The English LIWC contains a total of 76 different (sub)categories, while the German one contains 68. To perform sentiment analysis only the categories 'Posemo' (positive emotion; in the German version 'Posfeel' is also added) and 'Negemo' are relevant. The tool takes a tweet and parses it for words that are contained in the dictionary and assigns them to a category. There is no difference in intensity of sentiment of words and every word accounts for one point for each category it is assigned to. Then the relative scores to the length of the tweet (in words) are calculated and the 'Negemo' score of a tweet is subtracted from the 'Posemo' score. The result is then weighted by a scheme later described, as it is also applied for VADER. A tweet is then considered positive, if its score is greater than zero, neutral if it equals zero and negative if less than zero, so that it becomes a tertiary classification task.

[7]https://github.com/chbrown/liwc-python

#### *2.1.2 VADER.*

VADER is a dictionary based approach especially developed to adapt to characteristics specific to social media content. It expands the idea of a classical dictionary approach by a simple rule-based model. It uses lexical features and rates the content of a text accordingly with a sentiment intensity (hence valence based) score and then applies five rules that should reflect the ways social media users emphasize sentiment in their texts:

- - Exclamation marks increase the intensity of the sentiment (without changing its orientation)-> 'Working conditions are bad' > 'Working conditions are bad!!!'
- - Capitalization of whole words also increases the intensity -> 'Working conditions are bad' > 'Working conditions are BAD'
- - So called degree modifiers reduce or increase intensity based on how the modifiers are classified -> 'Working conditions are very bad' < 'Working conditions are bad' < 'Working conditions are a bit bad'
- - The word 'but' leads to a shift in polarity, and the overall score is dominated by the part after the 'but' -> 'Working conditions are bad, but I love my job' has a positive and a negative sentiment, but the overall score is dictated by the positive second half.
- - Negation changes the orientation of a sentence -> 'Working conditions are bad' < 'Working conditions aren't bad'

There is also a Python implementation available that already includes the necessary dictionary[8]. The application can also translate UTF-8 encoded emojis, identify slang like abbreviations like 'kinda' or decode typical acronyms like 'lol' that are relevant for the sentiment. It is designed for the English language and takes a tweet as input to generate four kinds of scores: a positive, a negative and, opposed to LIWC, a neutral score. The last score is the compound score, which is used to classify the tweets. The original VADER paper suggests using a threshold of <-0.05 to classify a text as negative, > 0.05 as positive and in between as neutral considering the compound score. Tymann et al. also created a German implementation of VADER [25], which also features a Python implementation[9], by using SentiWS as a starting point for the lexical features, adapting

[8]https://github.com/cjhutto/vaderSentiment
[9]https://github.com/KarstenAMF/GerVADER

language independent features like some heuristics and annotating additional relevant German opinion words. It is important to mention that the creator of the German implementation already mentioned some flaws of the application, as for example the negation is more complex in German and negation words like 'nicht' can appear with a higher distance to the word they negate and can also stand before or behind it. As negation is most of the time critical for the direction of the sentiment, especially the classification of German tweets will be reviewed in a more qualitative way after applying the dictionary based methods (LIWC does not account for negation at all). As VADER detects more nuances of tweets and also showed to outperform LIWC in various domains, a broader analysis is performed using it and LIWC is especially used to calculate the agreement between different methods to verify the findings, as no labelled data is available.

#### 2.1.3 *Weighting scheme.*

For the sentiment analysis part the decision is made to drop retweets entirely, the reasoning for which will be explained in the next section. The number of retweets is still a good indicator for how well a given tweet resonates with its readers. As there was no scheme present in the literature that incorporates the retweet score of a tweet into the sentiment, I use the following approach: since the popularity of a given user, as determined by the number of their followers, has a large influence on the retweet count, we normalize each tweet's retweet count by the user's follower count. The resulting number is an indicator of the popularity of a tweet relative to the popularity of the user. So less prominent users result in a higher influence, when they are retweeted a lot. To reduce the influence, this calculated ratio is normalized to a score between zero and one over all present tweets. Then the sentiment score is multiplied by 1 + the normalized weight-score. So the weighting scheme is the following with $r$ = retweet count, $f$ = follower count and $s$ = sentiment :

$$[\text{norm}(r/f) + 1] \cdot s$$

To avoid letting rather irrelevant tweets dominate this metric, all retweet-follower ratios for accounts with less than 50 followers are set to zero (the maximum retweet count below that threshold is 63 and this also reduces the influence of potential bots). As an example, the tweet in Figure 2 (all screenshots of tweets, with no explicit permission of use by the author are anonymized, to preserve the people's right to delete them) shows the highest ratio with 8.7. [H]


**Figure 2: Tweet with the highest retweet/follower ratio** [10]

One thing that has to be mentioned is that this weighting can not change polarity of the score at all for LIWC (neutral tweets are zero anyway), it can change neutral tweets for VADER, that are close to the threshold either to positive or negative tweets (depending on sign of compound score), but this is not the case for any tweets here. This scheme could also be applied to the likes of a tweet, but I chose retweets here, because they are a stronger indicator of support. This is the case, because retweeted tweets also appear in the timeline of one's followers.

## 2.2 Topic Modelling

In addition to the manual detection of topics, as well as the descriptive data analysis, topic modelling is used to identify further events and capture the content of conversations. Topic modelling itself is an unsupervised machine learning technique that is used to identify prevalent topics from unannotated, unstructured data. It does so by finding groups of items that are close together according to some distance metric. Furthermore it can be compared to clustering of numeric data for example, which is also used to find non-predefined classes. Especially for tweets it is useful to classify and group or sort them, so that it makes a possible future analysis and interpretation of single tweets feasible and directed. The approach can also lead to the discovery of hidden topics that are not obvious to a human observer. The relevant algorithms use the content of the text directly and have been applied to all kinds of documents and document collections, in domains like Twitter[14] or even fraud detection [30]. The goal is to identify repeated patterns in the words of various documents that results in them being similar and containing language about the same topic. Topics should mostly not be seen as a word like 'research', which is returned, but rather a set of words contained in the texts that belong together like 'precarious', 'conditions', 'research. By sorting the words that define a topic according to the frequency within the topic, it is then possible to interpret the topics using the most important words included in them. This makes it sometimes hard to label single topics with one or two terms, since lists of words can appear incoherent to a human reader. As a consequence I tried to make coherence measurable and the model fitted in this project will be optimized for this score mainly. This measure is used to ensure quality of single topics and improves the chances of making them interpretable. It is an evaluation method that started with using precision to determine coherence, but was then quickly abandoned and resurfaced only recently with several ways to measure it[6] and the exact score used will be explained after choosing the method used (as described below).

The paper of Churchill and Singh[6] evaluated a range of topic modelling techniques that are widely used and listed some of them like DMM (Dirichlet Multinomial Mixture) and LDA (Latent Dirichlet Allocation), which were the first models to be introduced in the early 2000s. The following models mostly aimed to improve LDA and overcome some of its flaws. With new types of texts and conversations occurring with the growth of social media, forums and blogs, other methods like NMF (Non-negative Matrix Factorization) or graph based algorithms were proposed. However adaptions of classical LDA still are state of the art for many domains. Zhao et

[10] https://bit.ly/3FNbjUc

al.[33] for example achieved good results for Twitter data with a Twitter specific adaption of LDA. The #IchBinHanna data will also be modelled by training a LDA model from scratch, as the overall topic itself is very specific and not very comparable to existing models or data, which tend to be to general, especially because conversation style on Twitter differs widely across communities.

LDA in a machine learning context was originally introduced by Blei, Ng and Jordan[3] and is a generative statistical model. It is used to generate topics and assign probabilities to documents that indicate how relevant they are related to the detected topics by their content. It assumes that there is a predefined number of topics (optimal number has to be determined), which are of course previously unknown (hence topics are 'latent'). Each of those follows a multinomial distribution over the whole vocabulary of the collection. Each document can be described by a distribution of topics (assumed to come from a Dirichlet distribution) and a topic is defined by a distribution of terms. LDA sees every text as a bag of words and topics are generated by co-occurrences of words in documents without taking their order into account. So the topics of a new document are determined by the words it includes and to which topic they belong. The Python implementation, which is used to build, apply and visualize the model is gensim [22].
As the data set is very large and the discussion is assumed to change over time, which means relevant topics can change and the temporal information is available, there are two ways to get relevant topics in general. The first one is to create a model that generates topics over the whole time span, which ignores the temporal dimension. To get more fine grained topics with this approach a larger amount of single topics has to be generated. But increasing the number of topics often reduces the coherence of single topics and increases the probability of overlapping words between topics. So the second approach, which I also follow in this project, is chosen to create a model for each month separately. This is furthermore a more beneficial approach, as there are clear indicators, like previously known events like the 'aktuelle Stunde' or the German federal election, for the change of topics over time. Also the amount of tweets posted in each month is very different and the dominance of the heavier months of June and July can be reduced with separate models. As a reference a model using the full data is also trained, but with a smaller amount of topics.
The next challenge is that LDA and topic modelling in general are usually not fit to detect topics for multilingual data. Although coherence won't be affected much, tweets of the same topic will usually not be assigned to the same topic, if their language is different. As the goal is rather to discover hidden topics and influential events and not to classify tweets per se, I decided to not split the models for every month into German and English. The main events driving the conversation are not expected to be different for the two languages, because the conversation is still largely evolving around a German law and events taking place in Germany.
Nevertheless, another model is built that only identifies topics of English tweets over the whole time, to confirm the assumption that topics overall are at least very similar across languages. To evaluate the quality of topics, coherence is measured like hinted before, along with perplexity, which the model is not optimized for though. The model is optimized for a coherence score introduced in the paper by Röder et al. [23]. It uses a one-segment sliding window of the top terms of a document and an indirect confirmation measure using normalized point-wise mutual information, which showed to align best with human interpretation in the paper, and cosine similarity. Perplexity basically describes how surprised a model is, when it's introduced new data. With the gensim implementation it is calculated as the normalized log-likelihood on a hold-out test set, but optimizing for perplexity often results in topics that are less interpretable for humans, as it has shown to often be not correlated to human judgement [4] or even be negatively correlated.

## 3 DATA ANALYSIS AND CLEANING

This section is subdivided into one part for each analysis approach, as the preparation of the initial data set has to be done differently for every approach. The original data was collected by a loosely connected group of researchers including the main supervisor of this project Dr. Jana Lasser and the code for the data collection is publicly available[11]. It consists of data sets for many hashtags that are related to the topic like #FristIstFrust, #IchBinReyhan, #WissZeitVG, which co-occurred with #IchBinHanna or discussed the issue before #IchBinHanna became viral. For this project, the subset of Tweets that was retrieved by querying Twitter for #IchBinHanna is used, which originally contains 116928 tweets from 19010 different authors. The subsection about the descriptive data analysis supports the detection of events. The insights gained there will later be combined with those of topic modelling.

### 3.1 Descriptive Analysis

As the movement deals with a German law and it originated in Germany, the majority of the tweets is in German (99219) as expected. There is still also a good amount of English tweets (13648), because academia is becoming more international and many affected researchers are not German native speakers and also because it gained attention across boarders. This leads to the first decision to only include German and English tweets, as the (dictionary based) sentiment analysis is language bound and topic models including languages that are heavily underrepresented do not hold much value either. There are 3147 tweets with undefined language and only 914 tweets with a language other than the selected ones.
The initial data set contains 93 columns (94 with the included user profiles), which were cut down to a set of 16 potentially relevant ones (author.description, author.id, author.name, author.public _metrics. followers_count, author.public_metrics.following_ count, author. username, author_id, created_at, id, lang, public_metrics. retweet_count, text, hashtags, reference_type, wanted_tag, user. group). The wanted tag here indicates whether the string #IchBinHanna (casing ignored) is actually contained in the tweet or not.

After taking a look at the data at hand, the next important step is to ensure the quality of the data. One major problem with the queried data set is that there are a lot of tweets that do not actually contain the desired hashtag. The problem is approached by first taking a qualitative look at the data and observing the tweets that do not directly contain the string #IchBinHanna (casing ignored). There are 52035 such tweets, from which 47468 are retweets. An

[11] https://github.com/LaserSteff/IchbinHanna

observed sample of those retweets gives an indication that those tweets are indeed related to the topic of other tweets containing #IchBinHanna. Some of them can even be verified to originally contain the hashtag. This surfaces the first problem: a lot of the retweets are cut off, because for example the addition of 'RT' plus the twitter handle at the start of a retweet leads to the tweet exceeding the maximum character limit. When looking at other tweets that are not flagged as retweet, the reason why they are included is less obvious. Some of them seem still related to the topic, but then there are a lot of tweets that seem absolutely arbitrarily included like:

> *'I finally heard from JW. Some of the misogynist & vilifying language has been removed if not all. It took some effort. Lasting damage has been done. We can defy it & yet the threat of defamation is real. No wonder it's mostly snr male w who write. The debate needs a new tone.'*

The presence of such tweets can not be explained by further looking at the data and maybe the original query could yield an answer, but for the analysis all tweets not containing the hashtag directly are removed. The remaining 60832 tweets, from which 40896 are retweets, include 52407 German and 4632 English tweets. Those tweets were posted by 13444 unique users, which results in a volume of about 4.5 relevant tweets per user.
After excluding the mentioned data, the remaining tweets are all relevant for this part of the analysis. The next goal is to identify events by analysing the raw tweets. If something related to the issue at hand happens, like the discussion of the topic during 'aktuelle Stunde' in the German Bundestag, one would expect people to talk about the topic to a greater extent, Therefore tweet frequencies for each day are calculated to detect peaks. To extract topics from those peaks, the most frequent words for at least the top three days per month will be investigated, in addition to days where obvious events (as determined by media coverage) happened that are not represented in a major spike in activity on Twitter. Also the blog[12] of Amrei Bahr, Sebastian Kubon and Kristin Eichhorn was a helpful resource to identify events without looking at the data directly.
The set is then further divided into a subset for each month, because the activity declined heavily, especially after July. While 37258 tweets were posted under the tag in June, in July there were still 13642, in August 5477 and in September only 4455. In order to also notice trends in months with a lower total output, they are observed separately. To get the mentioned top terms, the data has to be preprocessed. A first step is to remove URLs, as they hold no semantic information. Then to remove stop words that are also frequently used without adding to the meaning of a text, the tweets are lowercased and tokenized. The initial stop word lists for the removal are the German and English stop word dictionaries included in the nltk-pyhton package. After observing the top terms overall, some terms are added to the dictionaries (like 'innen' which is a leftover of gendering in the German language, when special characters are removed, or the string 'ichbinhanna' itself) and terms with two or fewer characters are dropped.

[12]https://ichbinhanna.wordpress.com

## 3.2 Sentiment Analysis

The preparation of the data for the remaining two approaches differs, as they both have their own needs in terms of data preprocessing. The tweets not containing the hashtag directly are also removed here. The next step is the same for both sentiment analysis and topic modeling, which is excluding retweets as well. For the sentiment analysis it is of interest to extract people's mood. While it can be assumed that people who retweet a tweet resonate at least a bit with the content of the original tweet, it does not express their sentiment directly. Furthermore the exact intention of a retweet can not always be determined. Also as the overall sentiment is of interest, this will reduce the influence of prominent accounts. To not exclude the information completely the introduced weighted score is used. After excluding retweets the data contains 19936 original tweets.

LIWC's dictionary contains only words that are assigned to a category. The English dictionary is slightly modified, as past works have shown that the word 'like' is of no particular use in the case of sentiment analysis, as its meaning is widely ambiguous in different contexts and it's quite highly frequented. Emojis, numerals, mentions and URLs are removed separately and then the tweets are lower cased and split into tokens using nltk's TweetTokenizer. From those, stop words are again removed. Tokens with 2 or less characters are also removed, which is an important step, because sentiment is calculated relative to tweet length. This step removes punctuation, as well as remaining tokens with little semantic meaning that the stop word dictionaries did not contain. No lemmatizing or stemming is used here, as LIWC detects various forms of the words in its dictionary.
For VADER a lot of those steps are not performed, as they would contradict some of the heuristic assumptions of the rule based system and the dictionary also accounts for different forms of most of its words. Punctuation is one important part, which is kept, because it is used to determine the strength of the given sentiment, as well as capitalisation. As VADER uses n-grams for negation for example, stop words are also not excluded. The only thing that is excluded, because it holds no information in this case are mentions and URLs.

## 3.3 Topic Modelling

For the topic models the set of tweets directly containing the tag and not containing retweets is used again. Retweets are excluded here, to avoid the generation of a single topic just because of one highly prominent tweet. The focus here lies on removing stop words to get meaningful topics that are not dominated by frequent words with low semantic value. URLs and emojis are also excluded for this analysis. Hashtags and mentions are kept, as some topics surface that reveal twitter profiles that had a relevant talking point often tied to an event. This is interesting for a following qualitative analysis of tweets in a topic.
The tweets are then again filtered using nltk's German and English stop word list and a list of individually selected words, tags and mentions that can be found in the topic modelling part of the github repository. The tags excluded include for example #IchBinHanna and #WissZeitVG. In addition, the mentions of the initiators of

the hashtag Amrei Bahr, Kristin Eichhorn and Sebastian Kubon are removed, because they stay relevant over the whole time span. After that, English and German tweets are lemmatized separately, to reduce the overall vocabulary. For English tweets nltk's WordNetLemmatizer [17] is used, which is very performant in general. However for German texts it often gives less satisfying results. The Hannover Tagger introduced in 2019 by Wartena[28] is therefore used, as it is a heuristic and hidden Markov model based approach developed specifically for German morphology.

## 4 RESULTS AND DISCUSSION

This section shows and discusses the results of the methods applied and problems that occurred. Both sentiment analysis and topic modelling are reported separately and then their results are put together, by mapping sentiment to the events detected.

### 4.1 Descriptive Data Analysis

With the previous knowledge about some events and the information of the descriptive analysis, the top terms of a total of 20 days (highest tweet volume per month or previously known events) are investigated. This leads to the detection of the following events:

- **10th of June**: Those spikes occurred at the start of the movement.
- **13th of June**: Answer from the Ministry of Education and Research[13], which got deleted from their website.
- **14th of June**: Initial video of the Ministry for Education and Research got deleted from their website.
- **17th of June**: State secretary Wolf-Dieter Lukas uploaded a video addressing the issues that arose with the hashtag[14].
- **24th of June**: The issue got discussed in the 'aktuelle Stunde' of the German Bundestag[15].
- **25th of June**: Still some tweets about the Bundestag, but also a series of tweets by Achim Landwehr, a dean of studies, went viral, in which he criticizes and portrays the situation of scientific workers from the perspective of 'Hannas' Boss' [16].
- **1st and 2nd of July**: Zoom conference of the GEW (a union for educators and researchers)[17].
- **9th of July**: Journalist Thilo Jung presents the issue at a press conference of the German government[18].
- **3rd of August**: German public-service broadcaster ZDF sends a report about the movement of #IchBinHanna on TV in the 'Heute Journal'[19].
- **2nd of September**: Declaration of the new 'Berliner Hochschulgesetz', a law that tries to ensure tenured contracts for scientific workers, if they meet their qualification goal in time[20].
- **26th of September**: German federal election.
- **29th of September**: #IchBinHanna was discussed in the Landtag of Hesse (parliament of German State of Hesse)[21].

The last part of the descriptive data analysis includes the background of the users reflected by the user profiles obtained. Note that for the main analysis (topic modelling and sentiment analysis) it does not matter what type of user sent a tweet, which means that all tweets that are declared as relevant in the following sections are kept. However, for the specific task of analysing user groups, the category 'Bots' is excluded, because it is not determined a relevant group in the academic scope, which the addition of user groups should reflect. I also decided to exclude teachers, junior professors and medical professors, because those groups sent less than 100 tweets over the whole time. The user profiles include a total of 13 categories, which are the following : 'Postdoc' (postdoctoral researcher), 'Academic unspecified' (related to an academic institution, but degree of education unknown), 'Promovierende' (doctoral researcher), 'Prof' (professor), 'Bot', 'Media' (person working for a media outlet or account of a media outlet), 'Student', 'Institution', 'Union rep' (representative of a union), 'Political rep' (representative of a political party), 'Jun. prof' (junior professor), 'Teacher', 'Medical doctor'. From the 13044 unique users, there are 5625 users that are classified, which account for a total of 40956 tweets. The distribution of the relevant classes concerning unique users can be found in Table 1 and the proportion of tweets of classified users over all user groups can be found in Table 2. It can be seen that

| Academic | Postdoc | Promov. | Prof | Media |
|---|---|---|---|---|
| 36.1% | 21.4% | 14.4% | 11.3% | 5.2% |
| Student | Political | Institution | Union | |
| 5% | 2% | 1.1% | 1% | |

**Table 1: Proportion of unique users.**

concerning users, about a third of the classified users could not be specified in detail, but are related to the academic system. Most of the users that are specifically classified are postdoctoral researchers, which makes sense, as this is the group that started the movement and the problem at hand is most relevant to them. When looking at the tweet volume per month, it can be observed that the proportion

[13] https://web.archive.org/web/20210613204441/https://www.bmbf.de/de/ichbinhanna—antwort-des-bmbf-auf-die-diskussion-in-den-sozialen-netzwerken-14675.html?s=09
[14] https://www.bmbf.de/bmbf/shareddocs/kurzmeldungen/de/ichbinhanna.html
[15] https://www.bundestag.de/dokumente/textarchiv/2021/kw25-de-aktuelle-stunde-jobs-wissenschaft-849096
[16] https://bit.ly/3qLw4LT
[17] https://www.gew.de/aktuelles/detailseite/ichbinhanna-fachtagung-twitter-bewegung-ruestet-sich-fuer-mehr/
[18] https://www.youtube.com/watch?v=OurzGBOS22E
[19] https://www.zdf.de/nachrichten/heute-journal/ich-bin-hanna-102.html
[20] https://www.tagesspiegel.de/wissen/gesetz-zur-staerkung-der-berliner-wissenschaft-was-sich-jetzt-fuer-berlins-universitaeten-und-fachhochschulen-aendert/27574344.html
[21] https://hessischer-landtag.de/termine/83-plenarsitzung

| Academic | Postdoc | Promov. | Prof | Media |
|---|---|---|---|---|
| 24.9% | 44.7% | 9.2% | 7.6% | 2.8% |
| Student | Political | Institution | Union | |
| 2.3% | 0.9% | 1.7% | 1.4% | |

**Table 2: Proportion of tweets across user groups.**

of the most prominent group of postdoctoral researchers grows over the months. While in the most prominent month of June, 26.7% of the classified tweets were sent by them, in July there were 33.2%, in August already 38.3% and in September the proportion is 39.2%. This is interesting, because the overall volume of tweets significantly drops over time and the discussion shifts more and more in the direction of directly affected persons.

## 4.2 Sentiment Analysis

As VADER is the tool that is more tailored towards social media data, more extensive analysis is performed on the results of this method. In general the expectation was that overall sentiment would be rather negative, as the purpose of the hashtag was often to share experience and to showcase one's situation within the German academic system. Such situations are often characterized by precarious working conditions and lacking career perspectives.

Starting with LIWC the first interesting finding is that for the analysis performed positive tweets account for the majority of tweets (ignoring neutral tweets). LIWC detects a total of 8888 tweets with a score of zero, which stands for neutral tweets, 7521 with a positive sentiment score, and only 3527 negative tweets(44.5%/37.8%/17.7%). The most positive score over the whole data is 0.67 and the most negative is -0.67 (for LIWC this is the proportion of positive words to the whole word count of a tweet minus the proportion of negative words and an added weight based on retweets) and the mean sentiment is 0.0249. To see if the results are consistent across both languages the counts of the German tweets are 7771 positive, 6506 neutral and 3037 negative (44.9%/37.6%,17.5%), while there were 1117 positive, 1015 neutral and 490 negative English tweets(42.6%/38.7%/18.7%). So both languages showed a quite similar distribution. The distribution of sentiment intensity can be seen in Figure 3 and it can be seen that the intensity of both classes is distributed around zero, which means that in most of the tweets none of the polarities is really dominant. When plotting the sentiment over time and aggregating the mean value per day, the dominance of positive over negative tweets becomes clear. Figure 4 shows the mean values of the whole time span and there are only two days with a negative mean sentiment. Those two days are the 11th of July with a score of -0.004 and the 25th of September with -0.002, so those are both only slightly negative. The findings from LIWC already differ from the initial intuition of having a rather negative sentiment overall.

When investigating the results of VADER, the direction of sentiment quickly is uncovered to be the same. VADER reports single scores for each sentiment polarity, as well as a neutral score and a compound score, that is a normalized value between zero and one to reflect overall sentiment. For the positive score the maximum value

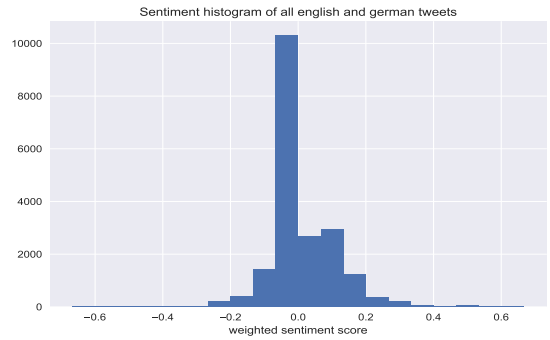


**Figure 3: Histogram of sentiment intensity for LIWC**

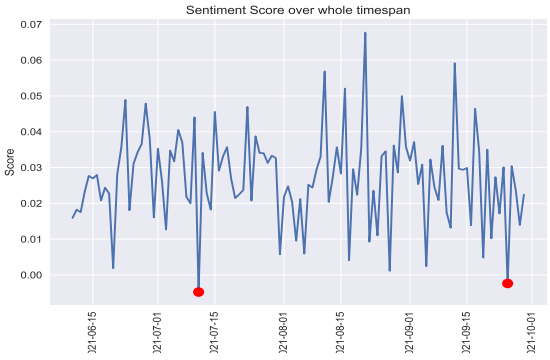


**Figure 4: Mean sentiment scores by day from June-September for LIWC**

is 0.86 and the mean is 0.11. For the negative score the maximum is 0.81 maximum and the mean is 0.05, and the neutral score has maximum of 1 and a mean of 0.83. The reported compound, which is the score to which weighting was applied, has a maximum of 1.01, a minimum of -0.99 and a mean of 0.21. For VADER positive tweets are even more frequent than neutral tweets. A side note here is that the weighting did not shift any neutral tweets into another class. VADER identified 11335 positive, 4415 negative and 4186 neutral tweets (56.9%/22.1%/21.0%). The sentiment is more distributed towards the extremes this times, which means that intensity varies more here, which is depicted in Figure 5. This is probably the case, because VADER has a real measure to calculated intensity, while it's just the proportion of polarity words for LIWC. This time it is even clearer that the overall sentiment is positive, as there is not a single day in Figure 6 that has a mean score below zero. The red and green line show the specified -0.05 and 0.05 threshold and there is not even a day that on average has negative or even neutral sentiment. The distribution across languages are fairly similar again, as there are 10045 positive, 3726 negative and 3543 neutral German tweets (58.0%/21.5%/20.5%) and 1290 positive, 689 negative and 643 neutral English tweets (49.2%/26.3%/24.5%). In Figure 7 we can see an interesting difference, as it shows that German VADER tweets have only positive mean sentiment scores for all days, while the

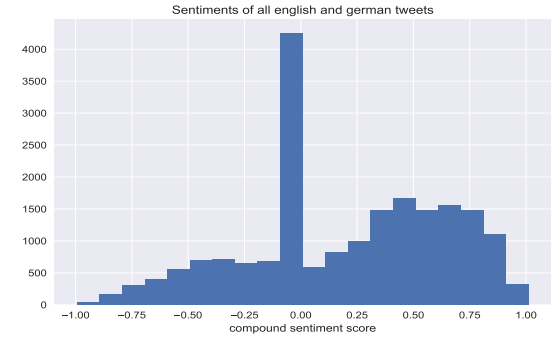


**Figure 5: Histogram of sentiment intensity for VADER**

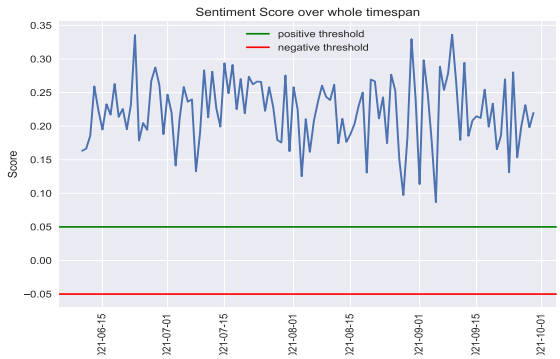


**Figure 6: Mean sentiment scores by day from June-September for VADER**

English data includes some neutral and negative days. It has to be taken into account that on the most positive day, there were only 4 English tweets and on the most negative one there were only 5.

The the overall positivity of the conversation was surprising and contradicts the assumption in the formulated research question. An explanation for this could be that the human language in general has a positivity bias[8], but there could be additional reasons too. Therefore analysis is further conducted on the content of positive and negative tweets for VADER, to get an impression of the quality of the results. The first property observed are the hashtags that co-occur in tweets containing #IchBinHanna, to see if those differ based on sentiment. The top four most frequent hashtags are the same for both positive and negative sentiment (#WissZeitVG, #IchBinReyhan, #HannaImBundestag, #95vsWissZeitVG) and the first one that differs is #FristIstFrust, which is ranked higher for positive tweets. This is interesting, because it has a rather negative connotation in German. Other than that there are no significant differences in hashtag co-occurrences.

Overall there are no hashtags that only occur in negative tweets, but there are some which do occur only in positive ones (e.g. #BTW21, #WasPostDocsWollen). Mostly this is due to the fact that there are more than twice as many positive tweets, but there are also clearly negative hashtags like #gruenermist in this list.

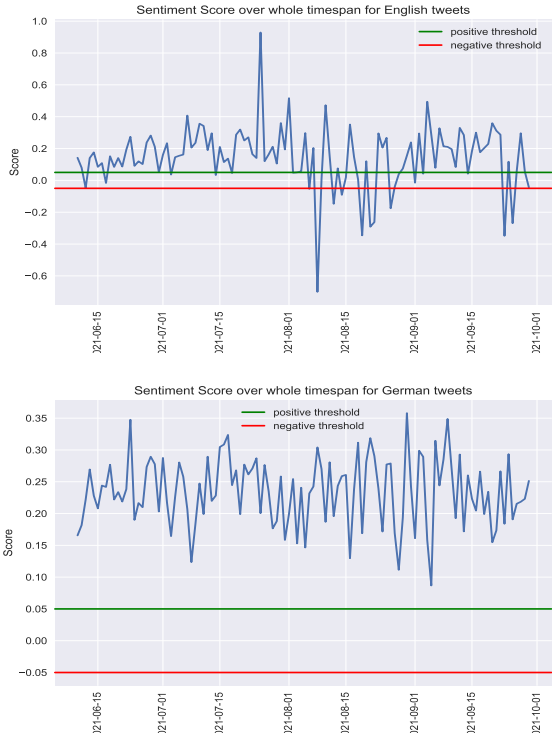


**Figure 7: Mean sentiment scores by day of English and German tweets for VADER**

Regarding content, hashtags are then removed from tweets to get word frequencies. For this the data is divided in six different parts, namely in German positive, German negative, German neutral, English positive, English negative and English neutral. For this procedure some additional custom stop words are excluded. This list of additional stop words includes actual stop words like 'us' or many and also names like 'Amrei Bahr' or terms like 'Wissenschaft' that were dominant within all classes. The word frequencies are visualized in the word clouds in Figure 8. The top shows German positive, negative and neutral tweets (from left to right and neutral below) and the bottom shows the same for English tweets. The words are quite similar for the positive and negative category, but some opinion words like 'danke' and 'gute' for positive German tweets and 'leider' or 'problem' for negative ones can be detected. For English tweets that was not so much the case, as words were very similar for both classes (also potential sentiment words like 'great', 'better', 'thanks' in negative classes).

As user groups are also of interest the average sentiment scores of each relevant group is shown in Table 3 for VADER and in 4 for LIWC. The mean sentiment values are relatively close to the mean for most of the groups, but there are some differences.Media has the lowest and political representatives have the highest mean sentiment score. To see if there are significant differences between groups a two-sample t-test for every combination of user groups is performed. At a significance level of 5% the sentiment of 'Postdocs' only differs from 'Academics unspecified' and 'Institutions'. The


**Figure 8: Word clouds of positive, negative and neutral tweets**

| Academic | Postdoc | Promov. | Prof | Media |
|---|---|---|---|---|
| 0.235 | 0.21 | 0.221 | 0.203 | 0.173 |
| Student | Political | Institution | Union | |
| 0.164 | 0.26 | 0.28 | 0.238 | |

**Table 3: Mean sentiment per group for VADER**

| Academic | Postdoc | Promov. | Prof | Media |
|---|---|---|---|---|
| 0.031 | 0.027 | 0.028 | 0.026 | 0.016 |
| Student | Political | Institution | Union | |
| 0.023 | 0.033 | 0.031 | 0.018 | |

**Table 4: Mean sentiment per group for LIWC**

same is the case for 'Profs'. For 'Promovierende' only 'Institutions' show a significantly different sentiment. As 'Postdocs', 'Profs' and 'Promovierende' account for more than 83% of the tweets of specified users ('Academic unspecified' excluded), it is interesting to see that the sentiment within those groups is similar (p-values >32% for the t-tests between those groups). The most differences can be seen for the groups of 'Media' and 'Students'. While they are very similar to each other (p-value 80%), they significantly differ from any other group than 'Postdocs' (p-value 10% and 10%) and 'Promovierende'(p-value 5% and 6%) with relatively low p-values and 'Profs'(p-value 18% and 21%). So concerning the formulated research question, it can be said that sentiment does not differ significantly within the most prominent groups, but there are differences with some of the smaller groups.

To make another reliability check, the agreement across VADER and LIWC results is calculated by using Cohens Kappa, which is at 0.35, on the sentiment labels. This is in general not a strong agreement, however 81.7% of the positive tweets of LIWC are also classified as positive for VADER, as well as 56.2% of the negative tweets, the biggest disagreement is on the side of the neutral tweets (37.1%). Because of the low overall agreement and the observed similarities from the content analysis, the tweets are also analysed qualitatively. This was done by manually analysing the content samples of both negative and positive tweets classified by VADER. After analysing 100 tweets of each class, the initial intuition of a rather negative sentiment can be observed in a a lot of positively labelled tweets.

When analysing the negative labeled tweets, I would agree on the majority of the labels and still classify them as negative. However this is not the case for positive tweets. When analysing them, the drawbacks of sentiment analysis techniques, especially dictionary based ones becomes obvious: A lot of the positive tweets could not be agreed on and based on the observed sample three types of misclassification could be identified multiple times. The first one contains obviously ironic or sarcastic tweets, like the example in Figure 9 (type 1). A characteristic of those tweets in the context of


**Figure 9: Example tweet containing irony**[22]

the topics, is that these people feel taunted by the positive display of Hanna's working conditions and respond by ironically talking positively about working overtime for example. The second type could be called mixed sentiment, but for the #IchBinHanna conversations this is a special case. There were a lot of tweets like the one shown in Figure 10 (type 2), where people talk about their

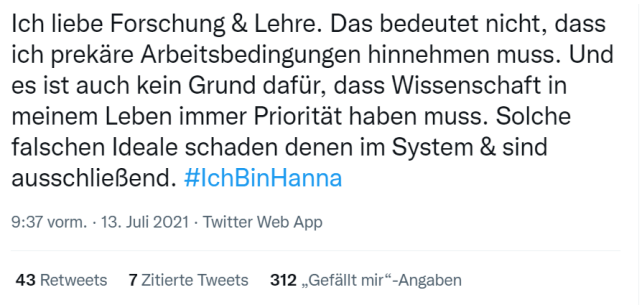


**Figure 10: Example tweet containing mixed sentiment**[23]

[22] https://bit.ly/3rLvNrc
[23] https://bit.ly/3oLSuLs

precarious working conditions in a very negative way, but in the same tweet they mention how they actually love what they are doing, which is working in as a scientist. Those tweets got a rather negative overall sentiment in most cases. The third type can be seen in Figure 11 (type 3), which clearly shows a negative tweet,

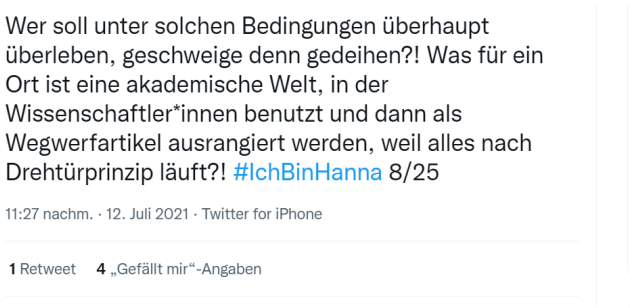


**Figure 11: Example of a negative tweet classified as positive**[24]

which was classified as positive. In this case words like 'Wegwerfartikel', 'ausrangiert' or 'Drehtürprinzip', which are important for the negativity of the tweet, are simply not contained in the dictionary of both tools. Also both sentences are rhetorical questions, which the methods can of course also not interpret. Another interesting insight, when qualitatively looking at the data, were the posts that were actually positive. A lot of those are tweets of people that thanked for or praised other tweets that contributed to the topic, or people talk positively about media coverage or expressed their gratefulness that the topic gained attention. For future work it would be interesting to manually annotate a set of the tweets and categorize them with a similar schema to get more interesting insights about the conversations, but detecting those is a difficult task on a word level alone.
This showed that the results of the dictionary based methods chosen here have to be interpreted carefully, but they both show a clear trend towards positive sentiment. However there are no reliable resources to evaluate them further as there are no available data annotations.

## 4.3 Topic Modelling

The topic modelling stage of the project involves various models created on the whole data and subsets of it and is performed using the gensim implementation of LDA for Python. The models created include one for the whole data set, one for English, one for German tweets and one model for the tweets of each month to get a more fine-grained insight. To get the best possible outcome the models are tuned using a grid search for the following parameters: the number of topics (2-10, with step size 1;this was readjusted manually if necessary), the number of passes (5-20, with step size 5), the alpha parameter (a = 'asymmetric', s = 'symmetric') and the decay (0.5-0.9, with step size 0.2). This grid search was used to optimize the coherence score and the obtained values are reported in Table 5. This already shows that creating single topic models for each month results in more coherent topics. For the visualization of the

[24]https://bit.ly/3fOwRVI

| | Full | English | German |
|---|---|---|---|
| topics | 6 | 7 | 8 |
| passes | 20 | 20 | 20 |
| alpha | a | a | s |
| decay | 0.5 | 0.5 | 0.5 |
| coherence | 0.29 | 0.23 | 0.18 |

| | June | July | August | September |
|---|---|---|---|---|
| topics | 7 | 7 | 7 | 6 |
| passes | 20 | 20 | 20 | 5 |
| alpha | a | a | a | a |
| decay | 0.5 | 0.5 | 0.5 | 0.9 |
| coherence | 0.31 | 0.35 | 0.44 | 0.43 |

**Table 5: Model parameters and coherence score for LDA**

topics, pyLDAvis[25] is used, which creates an intertopic distance map, on which each topic is visualized as a bubble with the number of documents within a topic encoded in the bubble size. The topics are arranged in a two dimensional space, in which their distance is shown. This two dimensional space is obtained by using multidimensional scaling on the inter topic distances, like explained by Chuang et al[5]. Next to the graph, there is another visualization that shows the top 30 most salient terms in the whole corpus per default. Whenever a topic is selected, the most relevant terms for the given topic are shown with a blue bar indicating the overall term frequency and a red bar showing the estimated frequency within this topic only.
Some of the topics show the shortcomings of the German and English stop word dictionaries of NLTK, which were even extended, but apparently still did not manage to remove all relevant stop words ('many' for example occurs in some of the English topics). An example for such an intertopic distance map can be seen in Figure 12. The visualization shows the topic distribution of the six

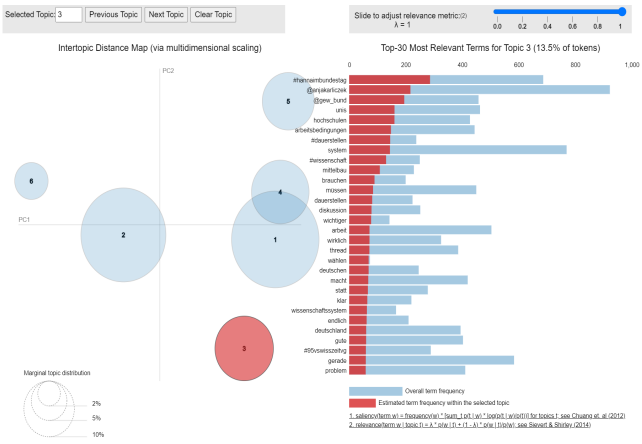


**Figure 12: Intertopic distance map for all tweets**

topics obtained by modelling the whole data. The visualized topic

[25]https://github.com/bmabey/pyLDAvis

is identified to be mainly about the previously known discussion of the 'aktuelle Stunde', where the issue was discussed in the German Bundestag. The second most important term in this topic is the Twitter handle of Anja Karliczek, the German secretary for Education and Research, who spoke about the topic at that event. Therefore, this topic clearly gives some valuable insight about an event discussed under the hashtag.
Before observing the other topics it is important to mention an interesting finding. Topic modelling for tweets in this context turned out to be a valuable source to identify users that were important in the discourse, either because they posted things relevant to the topic or like in the case of Anja Karliczek spoke publicly about it via other channels. Some events can be identified by inspecting why a certain Twitter handle appears as a salient term within a topic. Furthermore the handle appears with a context (other words in the topic), so that it can be derived why this mentioned profile is relevant in the discussion. This is also the case for hashtags, which often can sum up a topic or reflect an event directly (like #HannaImBundestag).
The output of each model were ten terms per topic that represent it (a whole list of all topics can be found in Appendix A) and samples of those are discussed here. To showcase the topic models, all topics of the full model are labelled by summing up their terms and the results can be seen in Table 6.

| Topic 1: Temporary contracts after promotion |
|---|
| jahren jahre stellen immer anjakarliczek warum promotion system einfach stelle |
| Topic 2: Discussion in the German Bundestag |
| #hannaimbundestag anjakarliczek heute arbeitsbedingungen #dauerstellen thema hochschulen letzten debatte unis |
| Topic 3: WissZeitVG and research in Germany |
| forschung lehre system deutschland gerade @gew_bund wisszeitvg #wissenschaft hochschulen problem |
| Topic 4: Labour Union GEW |
| system #hannabeidergew stelle @gew_bund forschung tweets @mahaelhissy #95wisszeitvg #acertaindegreeofflexibility macht |
| Topic 5: General topic about research in Germany (similar to topic 1) |
| deutschland forschung #ichbinreyhan jahre lehre bitte unis arbeit @gew_bund müssen |
| Topic 6: English topic about temporary contracts and working conditions |
| system research work years contracts working many conditions researchers time |

**Table 6: Topics for the whole data**

For the whole data set the main findings were that topics mostly discuss the issue at hand and they are each related to the overarching discussion under the #IchBinHanna tag. Topic 1 for example explicitly deals with the temporary contracts for scientific workers. Another common finding is that for the full model it can be seen that including both languages leads to a smaller topic that contains a lot of English words, however the overall theme about the video of Hanna and the problem of temporary contracts in scientific work is still coherently displayed here (as can be seen in topic 6). The scenario with small English topics can be seen for the other models using both languages as well, which is not considered an issue here, as those topics were still coherent. Another topic talks about #HannaImBundestag and includes Anja Karliczek, the GEW-Bund (a labour union), and the claims made by the scientific community that were the theme of the 'aktuelle Stunde' (example: #dauerstellen).
Interesting findings within the English model that supported the detection of events were the Twitter handles of mahaelhissy and kinofrau1, who brought up the topic of inclusion and tweeted in English about it to reach a more diverse audience. Other than that it mostly represented the general topics again and the same goes for the German model. Another important hashtag in the conversation that is reflected across topics is #IchBinReyhan, which raises the concern that the conversation has to be more inclusive and addresses issues of BIPOCs in academia.
For the model that takes only the tweets in June into account, dominant topics of course deal with Anja Karliczek and the 'aktuelle Stunde'. Also the general topics about precarious working conditions and the criticism of the system appear again. One topic for example most probably talks about how scientists dealing with humanities have problems to find jobs outside academia after temporary contracts expire.
In the month of July topics discover the media coverage in the 'Tagesthemen', the event of the labour union GEW that happened during that time and also interesting or relevant hashtags (#frististfrust, #acertaindegreeofflexibility, #tvstud for example). One topic in August includes a lot of Twitter handles, that are especially related to the series of tweets posted by Achim Landwehr, which were already identified as an event (other handles include karolinedoering, richterhedwig,esteinhauer).
The August model also revealed the involvement of ver.di, which is another German union. The September tweets included topics about the federal election, the 'Berliner Hochschulgesetz' and again claims that were made to politicians, especially because of the elections. September reveals a flaw of the approach of taking a single month, as it was the month with the least tweets.
The September model returns a topic that completely consists of terms and hashtags of one account, that posted the same tweet over ten times and is most likely a bot. This however encourages the approach to exclude retweets. Also the Twitter handle of the 'Netzwerk für Gute Arbeit in der Wissenschaft' (@NGAWiss) is found to be important in one of the topics, because they among others released a video as an answer to the initial Hanna video. The data from the topic models was very valuable to be able to go on with a directed search within tweets, news articles and other resources to identify events that were relevant in Twitter conversations. It was also helpful to plot the frequencies of each topic over time, to see at which days a specific topic was more relevant, which can be seen in Figure 13, which shows the frequencies in August (all frequency plots can be found in Appendix B).

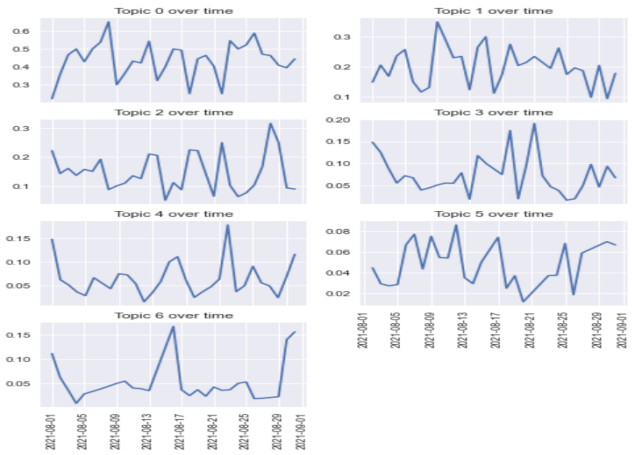


**Figure 13: Topic frequencies over time in August**

Clear spikes are visible for some topics, which then are further investigated.
The individual investigation lead to the identification of the following additional events that conclude the detection of the most important events and topics, which was among the goals of this project:

- **11th of June**: Series of highly interacted tweets by Dr. Hanin Hannouch, raising the issues of especially BIPOCs within academia[26]
- **3rd of July**: Demonstration #TVStud (initiative for labour agreements for student workers) organized by GEW and ver.di in Hannover [27]
- **8th of July**: Media coverage of #IchBinHanna from ARD's 'tagesthemen' featuring a contribution of initiators Amrei Bahr and Kristin Eichhorn, but also a heavily criticized comment afterwards [28] [29]
- **20th of August**: Satirical TV report of ZDF's 'heute-show' about the issue of #IchBinHanna [30]
- **26th of August**: Trending of #WasPostdocsWollen as a reaction to a statement of Günter Ziegler, president of FU Berlin, in 'der Tagesspiegel' [31]
- **1st of September**: Demonstration 'Entfristung jetzt' in Frankfurt (and other Hessian cities) as a reaction to the start of collective bargaining of public service in Hessia [32]
- **23rd of September**: Video of GEW, ver.di and 'Netzwerk für Gute Arbeit in der Wissenschaft' as an answer to the initial Hanna-video [33]

To answer the question which topics, specific user groups talk about, we have to take a look at distribution of topics within each group. Table 7 shows the ratio of the number of tweets about a topic to the total amount of tweets of a user group. We can already see the dominance of topics 1 and 2 here, which have the most tweets for all user groups. This is further clear, when looking at the histogram in Figure 14. Those two topics talk about the temporary contracts, which is the main theme of the whole movement and the discussion in the 'Bundestag', which happened in the month of June, where the tweet volume was the highest. So it makes sense that the whole conversation focused around those topics. However there are groups that prioritized some topics more than others or did not talk much about another topic. 20% of the Media tweets for example are about the WissZeitVG topic. This topic is more about the notions of the system that cause the problems. As the media is not affected by the problems, it can be assumed that they report more about the system itself and the law that causes problems for researchers. A topic that is less represented in the media tweets than others is the topic about the labour union GEW. This topic appears in no tweet of a political representative, which is also an immediate observation.

| | Academic | Postdoc | Promov. | Prof | Media |
|---|---|---|---|---|---|
| Topic 1 | 39.7% | 38.5% | 40.2% | 40.1% | 36.2% |
| Topic 2 | 24.4% | 25.8% | 23.1% | 25.7% | 30.7% |
| Topic 3 | 14.3% | 11.2% | 12.8% | 11.3% | 20.4% |
| Topic 4 | 11.7% | 15.5% | 15.5% | 14.9% | 2% |
| Topic 5 | 5.8% | 5.8% | 5.4% | 4.5% | 6.7% |
| Topic 6 | 4% | 3.2% | 3% | 3.5% | 4% |

| | Student | Political | Institution | Union |
|---|---|---|---|---|
| Topic 1 | 35.7% | 48.9% | 51.3% | 53.2% |
| Topic 2 | 34.9% | 27% | 18.9% | 16.7% |
| Topic 3 | 9.5% | 14.9% | 10.9% | 5.4% |
| Topic 4 | 9.5% | 0% | 7.2% | 1.6% |
| Topic 5 | 8.3% | 7.9% | 7.2% | 19.9% |
| Topic 6 | 2% | 1.4% | 4.5% | 1.6% |

**Table 7: Topic distribution of user groups for the whole model**

[26] https://bit.ly/3ArpYmR
[27] https://www.gew.de/aktuelles/detailseite/aktionstag-studentische-hilfskraefte-aufbruchstimmung-erzeugen
[28] https://www.youtube.com/watch?v=H1wJmqpGhJc&t=1s
[29] https://twitter.com/tagesthemen/status/1413239263227961344
[30] https://www.youtube.com/watch?v=5aVDRdQeBZM
[31] https://www.tagesspiegel.de/wissen/tenure-track-fuer-alle-nach-der-promotion-streit-um-dauerstellen-fuer-postdocs-in-berlin/27549238.html
[32] https://www.faz.net/aktuell/rhein-main/region-und-hessen/gewerkschaft-fordert-mehr-dauerstellen-an-hochschulen-17512258.html
[33] https://www.youtube.com/watch?v=XUt-48SIzOc

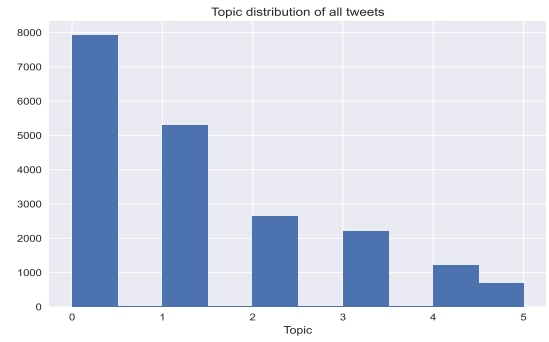


**Figure 14: Histogram of topics for the whole model**

## 4.4 Sentiment and Events

Bringing together the findings of the last two sections, this chapter shows the sentiment at specific events, to answer the final research question about the sentiment at specific events. As the obtained sentiment scores are overall positive, the results of VADER are used here, because they give an information about actual polarity. The goal is to get an impression on how the sentiment of the general conversation was when talking about a specific event (as far as results are reliable). With an exception, all the events discovered in the data analysis and the topic modelling stages, are reported here, as they showed clear indicators to be dominant in the conversation at that specific point in time. So by this dominance one can expect that the sentiment score is to a great extent influenced by the discussion and reception of these events when they happened, but not exclusively. The events can be seen in Figure 15 to Figure 18 sorted by month to get a better resolution on the time dimension. On 11th the highly interacted tweet series by kinofrau1 is not included, as it fell into the time the movement started, and the overall tweet frequency is too high to capture the sentiment of this event.

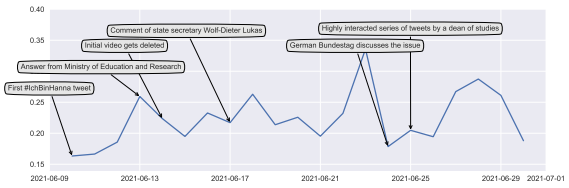


**Figure 15: Sentiment and events in June**

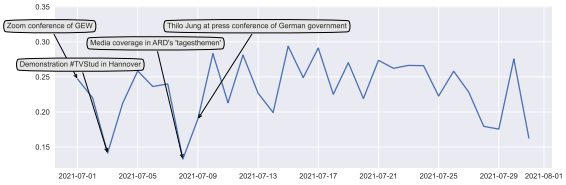


**Figure 16: Sentiment and events in July**

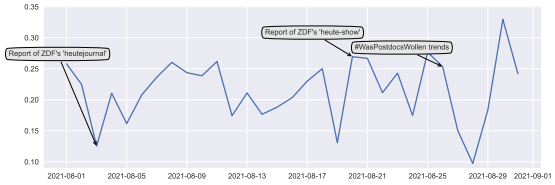


**Figure 17: Sentiment and events in August**

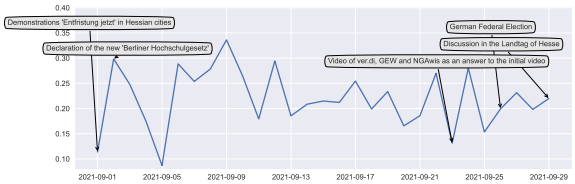


**Figure 18: Sentiment and events in September**

In June it can be seen that the movement started with an global minimum for June and increases towards the reaction of the Ministry of Education and Research, where a peak occurs. This is surprising, as the reaction was not very well received and got since deleted from the website. Shortly after that, when the video that caused the outbreak of the movement got deleted, the sentiment declines. The maximum sentiment score was reached just one day, before the issue got discussed in the Bundestag. At the day of the discussion the sentiment has a steep drop. A possible reason for that could be that the movement was positive that the issue got enough attention to be discussed at that level, but they were not so content with the discussion itself.

Looking at the development in July, the global minimum can be seen on the day, where ARD's 'tagesthemen' covered the issue. Although there was a contribution of initiator Amrei Bahr and Kristin Eichhorn, the tweets under the hashtag heavily criticized a comment by the 'tagesthemen' that followed afterwards and is most likely responsible for the low sentiment. Another low appears a few days before, when a demonstration took place in Hannover to demand fair wages, unlimited contracts etc. On this opportunity many users again raised all their concerns about the current state of the system, which would be expected to have a rather negative tone. Also the sentiment was a lot higher the days before, when the

labour union GEW hosted a Zoom conference, with a constructive discussion about the topic.

The three detected events in August all show a different reception. First another report of a public broadcaster is found at a day with a low sentiment score. However another format by the same broadcaster ZDF later in the month has got a more positive reception in terms of sentiment. This format was the 'heute show', which is a satirical TV-show. It heavily criticized Anja Karliczek and the working conditions the WissZeitVG creates. This is an interesting difference, because the more neutral report was discussed on a day with a sentiment score of 0.13, while the day with the satirical report that took the perspective of the affected researchers, has a mean sentiment of 0.27. The third event of the series of tweets under the hashtag #WasPostDocsWollen not only affected one day. On the 26th of August, the day the hashtag first occured, there were 17 tweets about, while the following day there were 27. From the 26th the sentiment dropped heavily, so it can be assumed that the discussion under that hashtag is less positive. This is reasonable, because it was again an aggregation of tweets about issues and demands of researchers.

Among the noticeable observations in September is the low sentiment at the first day, where demonstrations in various Hessian cities took place. For the demonstration in July, we could already also see a local minimum of the sentiment. Following that the sentiment rose again on the day the declaration of the new 'Berliner Hochschulgesetz' took place, but dropped to the lowest overall sentiment shortly after. The last note is about the video created by a group of labour unions that should serve as an answer to the original Hanna video. On that day the sentiment is surprisingly low again, as the video itself reflects the opinion of the people involved in the movement.

## 5 CONCLUSION

Analysing the discussion of #IchBinHanna lead to many interesting findings. While the overall positive sentiment was surprising, it has to be carefully interpreted. The general positivity bias, as well as the types of misclassified tweets are among the reasons for that. The qualitative analysis showed that a lot of the tweets express that researchers feel taunted by the laws and politicians and their reactions. This leads to a prevalence of irony and sarcasm in a lot of these tweets. This can not be detected by the chosen dictionary based tools and future research could focus on this. For both the topic modelling and the sentiment analysis it was beneficial to exclude retweets, which was especially important for the former, as tweets with the exact content can dominate or create a whole topic. Many applications of topic modelling on Twitter exclude hashtags and mentions as a pre-processing step. However in this case it turned out to be important to keep them, because whole topics revolved around single persons and hashtags were a good indicator for concrete events. The combination of descriptive analytics and topic modelling resulted in a list of many events that shaped the whole discussion and which can be further qualitatively analyzed. The majority of different user groups showed similar sentiment and tweeted about similar topics. As a lot of the discovered topics were quite general, it would be interesting for a further analysis to look qualitatively look at the content of different user groups specifically.

## 6 SELECTED DOMAIN-SPECIFIC LECTURE

Prior to starting the project, the lecture 'Computational Social Simulation'[34] related to the field of Computational Social Science was successfully attended in the 2021 summer semester.

## REFERENCES


[1] Himanshu Batra, Narinder Singh Punn, Sanjay Kumar Sonbhadra, and Sonali Agarwal. 2021. BERT-Based Sentiment Analysis: A Software Engineering Perspective. *Database and Expert Systems Applications* (2021), 138–148. https://doi.org/10.1007/978-3-030-86472-9_13

[2] Christos Baziotis, Athanasiou Nikolaos, Alexandra Chronopoulou, Athanasia Kolovou, Georgios Paraskevopoulos, Nikolaos Ellinas, Shrikanth Narayanan, and Alexandros Potamianos. 2018. NTUA-SLP at SemEval-2018 Task 1: Predicting Affective Content in Tweets with Deep Attentive RNNs and Transfer Learning. 245–255. https://doi.org/10.18653/v1/S18-1037

[3] David Blei, Andrew Ng, and Michael Jordan. 2002. Latent Dirichlet Allocation. In *Advances in Neural Information Processing Systems*, T. Dietterich, S. Becker, and Z. Ghahramani (Eds.), Vol. 14. MIT Press. https://proceedings.neurips.cc/paper/2001/file/296472c9542ad4d4788d543508116cbc-Paper.pdf

[4] Jonathan Chang, Jordan Boyd-Graber, Sean Gerrish, Chong Wang, and David Blei. 2009. Reading Tea Leaves: How Humans Interpret Topic Models. *Neural Information Processing Systems* 32, 288–296.

[5] Jason Chuang, Daniel Ramage, Christopher Manning, and Jeffrey Heer. 2012. Interpretation and trust: Designing model-driven visualizations for text analysis. *Conference on Human Factors in Computing Systems - Proceedings* (05 2012). https://doi.org/10.1145/2207676.2207738

[6] Rob Churchill and Lisa Singh. 2021. The Evolution of Topic Modeling. *ACM Comput. Surv.* (dec 2021). https://doi.org/10.1145/3507900 Just Accepted.

[7] P. S. Dandannavar, S. R. Mangalwede, and S. B. Deshpande. 2020. Emoticons and Their Effects on Sentiment Analysis of Twitter Data. In *EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing*, Anandakumar Haldorai, Arulmurugan Ramu, Sudha Mohanram, and Chow Chee Onn (Eds.). Springer International Publishing, Cham, 191–201.

[8] Peter Sheridan Dodds, Eric M. Clark, Suma Desu, Morgan R. Frank, Andrew J. Reagan, Jake Ryland Williams, Lewis Mitchell, Kameron Decker Harris, Isabel M. Kloumann, James P. Bagrow, Karine Megerdoomian, Matthew T. McMahon, Brian F. Tivnan, and Christopher M. Danforth. 2015. Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences* 112, 8 (2015), 2389–2394. https://doi.org/10.1073/pnas.1411678112 arXiv:https://www.pnas.org/content/112/8/2389.full.pdf

[9] Andrea Esuli and Fabrizio Sebastiani. 2006. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining.

[10] Namrata Godbole, Manjunath Srinivasaiah, and Steven Skiena. 2007. Large-Scale Sentiment Analysis for News and Blogs. *ICWSM 2007 - International Conference on Weblogs and Social Media.*

[11] C.J. Hutto and Eric Gilbert. 2015. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014.*

[12] Konsortium Bundesbericht Wissenschaftlicher Nachwuchs. 2021. Bundesbericht Wissenschaftlicher Nachwuchs 2021. buwin.de

[13] Sebastian Kubon. 2021. *First #IchBinHanna tweet by Sebastian Kubon.* https://twitter.com/SebastianKubon/status/1402886172158873600

[14] Jey Lau, Nigel Collier, and Timothy Baldwin. 2012. On-line Trend Analysis with Topic Models: twitter Trends Detection Topic Model Online. 1519–1534.

[15] Ji Yeon Lim, Jae Yoel Yoon, Lee Joon Kim, and Ung Mo Kim. 2012. Information Extraction of Review Using LIWC. *International Journal of Future Computer and Communication* 1, 2 (2012), 91.

[16] Prem Melville, Wojciech Gryc, and Richard Lawrence. 2009. Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09*, 1275–1284. https://doi.org/10.1145/1557019.1557156

[17] George Miller. 1995. WordNet: A Lexical Database for English. *Commun. ACM* 38 (11 1995), 39–. https://doi.org/10.1145/219717.219748

[18] Eva Murasov. 2021. *IchBinHanna trendet auf Twitter.* https://www.tagesspiegel.de/wissen/aufschrei-des-wissenschaftlichen-nachwuchses-ichbinhanna-trendet-auf-twitter/27278532.html

[19] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification Using Machine Learning Techniques. *EMNLP* 10 (06 2002). https://doi.org/10.3115/1118693.1118704


[34] https://bit.ly/3qFMYeC

[20] James Pennebaker, Martha Francis, and Roger Booth. 1999. Linguistic inquiry and word count (LIWC). (01 1999).

[21] Tanasanee Phienthrakul, Boonserm Kijsirikul, Hiroya Takamura, and Manabu Okumura. 2009. Sentiment Classification with Support Vector Machines and Multiple Kernel Functions. In *Neural Information Processing*, Chi Sing Leung, Minho Lee, and Jonathan H. Chan (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 583–592.

[22] Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50. http://is.muni.cz/publication/884893/en.

[23] Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the Space of Topic Coherence Measures. *WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining* (02 2015), 399–408. https://doi.org/10.1145/2684822.2685324

[24] Andranik Tumasjan, Timm Sprenger, Philipp Sandner, and Isabell Welpe. 2010. Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *Proceedings of the International AAAI Conference on Web and Social Media* 4, 1 (May 2010), 178–185. https://ojs.aaai.org/index.php/ICWSM/article/view/14009

[25] Karsten Tymann, Matthias Lutz, Patrick Palsbröker, and Carsten Gips. 2019. GerVADER -A German adaptation of the VADER sentiment analysis tool for social media texts.

[26] Bhagyashri Wagh, J. V. Shinde, and Nisha R. Wankhade. 2016. Sentimental Analysis on Twitter Data using Naive Bayes. *International Journal of Advanced Research in Computer and Communication Engineering* 5 (2016), 316–319.

[27] Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit Sheth. 2012. Harnessing Twitter 'Big Data' for Automatic Emotion Identification. https://doi.org/10.1109/SocialCom-PASSAT.2012.119

[28] Christian Wartena. 2019. A Probabilistic Morphology Model for German Lemmatization.

[29] Christine Westerhaus. 2021. *Twitter-Aktion ichbinhannaProtest / gegen Zeitvertragsgesetz flammt erneut auf.* https://www.deutschlandfunk.de/twitter-aktion-ichbinhanna-protest-gegen-zeitvertragsgesetz-100.html

[30] Dongshan Xing and Mark Girolami. 2007. Employing latent dirichlet allocation for fraud detection in telecommunications. *Pattern Recognition Letters* 28, 13 (2007), 1727–1734.

[31] Ronghua Xu and Qingpeng Zhang. 2016. Understanding Online Health Groups for Depression: Social Network and Linguistic Perspectives. *Journal of Medical Internet Research* 18 (03 2016), e63. https://doi.org/10.2196/jmir.5042

[32] Martin Zeyn. 2021. *WIESO DIE UNIS PERSPEKTIVLOSIGKEIT FÖRDERN.* https://www.br.de/kultur/ichbinhanna-ausbeutung-akademiker-sicherheit-uni-jobs-befristet-prekariat100.html

[33] Wayne Zhao, Jing Jiang, Js Weng, Jing He, Ee-Peng Lim, Hongfei Yan, and Xiaoming Li. 2011. Comparing Twitter and Traditional Media Using Topic Models. *Advances in Information Retrieval* 6611/2011, 338–349. https://doi.org/10.1007/978-3-642-20161-5_34

# A RESULTS OF ALL TOPIC MODELS

| Topic 1 |
|---|
| system contracts many work conditions research working employment precarious temporary |
| Topic 2 |
| scholar system @mahaelhissy @kinofrau1 acdemics @gew_bund need much time @akellergew |
| Topic 3 |
| years permanent research time work every contract postdoc career position |
| Topic 4 |
| work without contract event health @anjakarliczek research position scholars movement |
| Topic 5 |
| thread #ichbinreyhan system great today english better summary work academics |
| Topic 6 |
| story system share want researchers permanent experiences #ichbinreyhan well good |
| Topic 7 |
| career working researchers conditions research many system like people time |

**Table 8: Topic models for English tweets**

| Topic 1 |
|---|
| jahre promotion @anjakarliczek stelle zeit immer thema beitrag hochschulen gemacht |
| Topic 2 |
| #hannaimbundestag @anjakarliczek immer frage danke frau heute #karliczek bundestag einfach |
| Topic 3 |
| #ichbinreyhan @anjakarliczek #frististfrust gute heute jahre danke #95vswisszeitvg @gew_bund jahren |
| Topic 4 |
| @gew_bund #dauerstellen wirklich jahren deutschland macht besser #ichbinreyhan immer daueraufgaben |
| Topic 5 |
| stellen gerade immer stelle warum jahre eigentlich unbefristete unis befristete |
| Topic 6 |
| hashtag letzten forschung stunden problem dabei stellen danke unis arbeit |
| Topic 7 |
| forschung lehre müssen arbeitsbedingungen arbeiten menschen system statt arbeit heute |
| Topic 8 |
| jahre forschung lehre deutschland system arbeit dafür jahren dank promotion |

**Table 9: Topic models for German tweets**

| Topic 1 |
| --- |
| jahren immer system jahre @anjakarliczek thread #hannaimbundestag forschung warum letzten |
| Topic 2 |
| #hannaimbundestag @anjakarliczek forschung jahre lehre arbeitsbedingungen gerade heute hochschulen system |
| Topic 3 |
| @anjakarliczek gerade @gew_bund #hannaimbundestag menschen hätte müssen stellen zeit stelle |
| Topic 4 |
| promotion schreiben problem stellen stunde sichere jahre heute arbeitsbedingungen thema |
| Topic 5 |
| without karliczek passports scholars tweets conditions working einfach #hannaimbundestag |
| Topic 6 |
| diskussion bmbf antwort #hannaimbundestag evaluation anschlussverwendung @anjakarliczek zusammenfassung wisszeitvg forsche |
| Topic 7 |
| contracts research years career many system people researchers permanent precarious |

**Table 10: Topic models for June**

| Topic 1 |
| --- |
| #ichbinreyhan @anjakarliczek @tagesthemen gerade lehre jahren danke @gew_bund forschung vielen |
| Topic 2 |
| system #ichbinreyhan müssen #hannabeidergew genau arbeit dafür dabei einfach niemand |
| Topic 3 |
| work research #hannabeidergew time scholars working conditions teaching like contracts |
| Topic 4 |
| immer story einfach gerade urlaub denen share daran woche akademische |
| Topic 5 |
| @anjakarliczek heute deutschland #ichbinreyhan warum natürliche danke gute forschung debatte |
| Topic 6<br>stellen immer problem arbeit arbeitsbedingungen berliner @gew_bund befristete hochschulleitungen #ichbinreyhan |
| Topic 7 |
| #keineausnahme @anjakarliczek bestimmt #tvstud @niconolden beschäftigen märchen erfindung zeit beitrag |

**Table 11: Topic models for July**

| Topic 1 |
| --- |
| #ichbinreyhan system heute dafür zeit arbeiten arbeitsbedingungen immer einfach #waspostdocswollen |
| Topic 2 |
| #ichbinreyhan @anjakarliczek system forschung zeit unis #frististfrust eher macht situation |
| Topic 3 |
| immer beitrag @akellergew @karolinedoering thread @richterhedwig @esteinhauer @tinido @klios_spiegel @achimlandwehr |
| Topic 4 |
| stellen gerade jahren #dauerstellen immer lben angst heute frage forschung |
| Topic 5 |
| #waspostdocswollen @gew_bund wirklich stelle @jmwiarda vertrag #dauerstellen statt zwei vielleicht |
| Topic 6 |
| permanent #wirsindhanna power #gendergaga many postdoc change years scientists positions |
| Topic 7 |
| arbeit bleibt hochschulen danke thema bedingungen politik jahn druck @anjakarliczek |

**Table 12: Topic models for August**

| Topic 1 |
|---|
| #ichbinreyhan postdocs heute #hannainzahlen @gew_bund gerade stelle #entfristethanna promotion hochschulen |
| Topic 2 |
| #ichbinreyhan @gew_bund #dauerstellen thema #berlhg immer neue arbeit berlin bleiben |
| Topic 3 |
| @gew_bund immer stellen macht #ichbinreyhan heute warum jahren eigentlich dank |
| Topic 4 |
| @andreasbovensc @swh_hb @janinabruenjes #hannawählt-erinnerungstweet @spdlandbremen wählen #ichbinreyhan #r2g #universität #afd |
| Topic 5 |
| #ichbinreyhan @gew_bund immer @jenniferhenkehb #dauerstellen thread endlich hochschulen wegen @nga_wiss |
| Topic 6 |
| jahre davon kinder lehre system forschung arbeitsverträge change statt @gew_bund |

**Table 13: Topic models for September**

## B ALL TOPIC FREQUENCIES

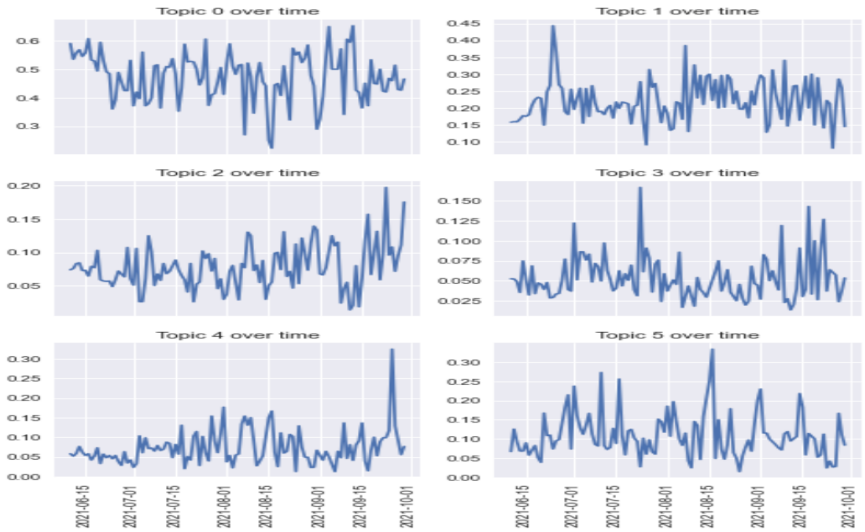


**Figure 19: Topic frequencies for the whole data**


**Figure 20: Topic frequencies for the German data**

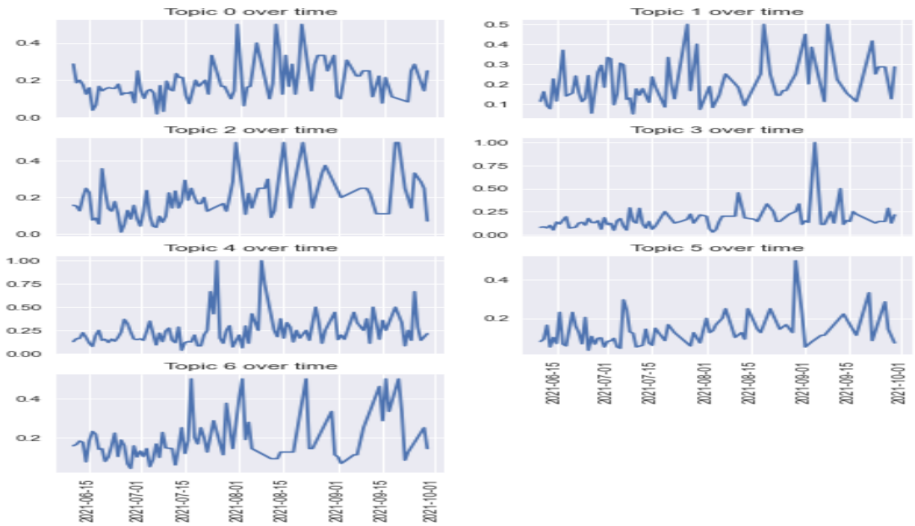


**Figure 21: Topic frequencies for the English data**

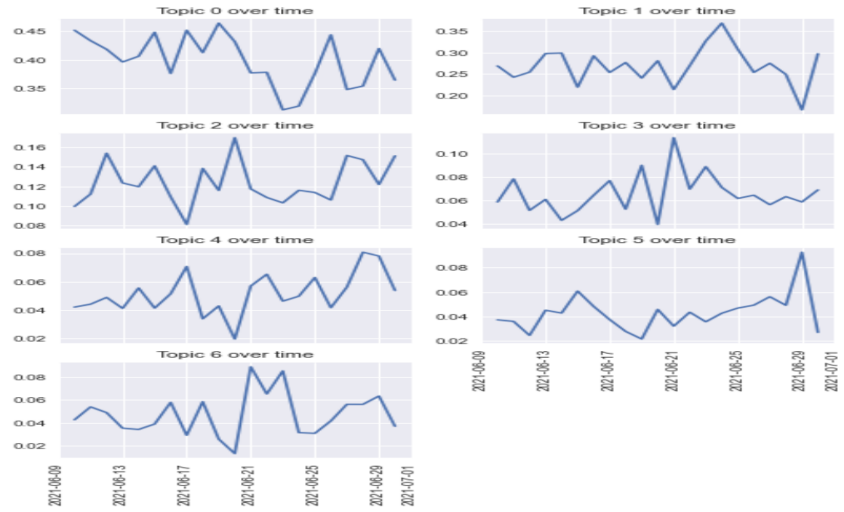


**Figure 22: Topic frequencies for June**

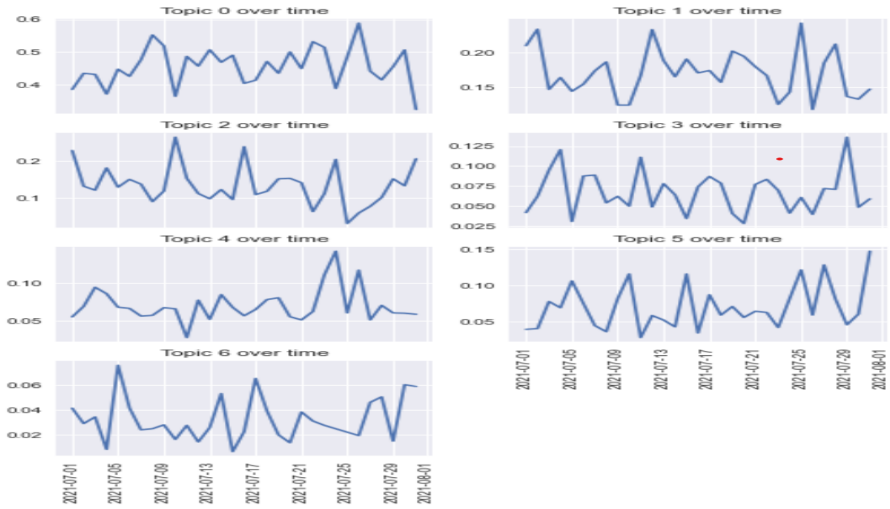


**Figure 23: Topic frequencies for July**


**Figure 24: Topic frequencies for September**