Political Debates in Parliament and on Twitter: Studying German MPs

Lawrence Fulton, Kristin Gnadt, Jona Rekate

7th March, 2024

1 Introduction

Twitter (now known as X)[1] is a prime example of how social media and privately owned companies impact political landscapes. Twitter usage influences politicians' success in elections (Kruikemeier, 2014; Udanor, Aneke, & Ogbuokiri, 2016) and there has been an increasing polarisation on Twitter (Garimella & Weber, 2017). Twitter is a way for politicians to cater to a bigger audience, to share their personal views and to promote their or their party's political goals (Bright et al., 2020). In order to better understand politicians' use of Twitter, we wanted to find out which topics are being discussed in tweets by members of the German parliament (MPs). We compare these findings to the topics of parliamentary speeches to figure out whether MPs behave differently on Twitter and in parliament. We therefore formulated the following research question:

RQ: Do the debated topics on Twitter by German politicians differ from the topics debated in the German parliament?

In order to answer this question, we use BERTopic, a topic modelling approach, to extract the most discussed topics on Twitter and parliament by German MPs in the 19th legislation period (2017-2021). We then compare the resulting topics of both datasets by computing their rank correlations. Additionally, we categorise the topic representations regarding their wider political theme and also compute those rank correlations.

In the past, it has been extensively researched how politicians in various countries and contexts use Twitter (Stier, Bleier, Lietz, & Strohmaier, 2020; Schürmann & Stier, 2023; Silva & Proksch, 2022; Golbeck, Grimes, & Rogers, 2010; Kreiss, 2016), and the speeches in many parliaments have been analysed regarding various items and with varying methods and scopes (Frid-Nielsen, 2018; Benoit, Schwarz, & Traber, 2012; Geese, 2020; Marcinkiewicz & Stegmaier, 2019).

However, to our knowledge, there is no study applying topic modelling to both Twitter and parliament and comparing the topic distributions, especially in the German context.

In recent years, topic modelling has become an important method to deal with large quantities of textual data (Zhao et al., 2011) and has been applied to both Twitter and parliament data (Curran, Higham, Ortiz, & Vasques Filho, 2018; Zhao et al., 2011; Contreras, Verbel, Sanchez, & Sanchez-Galan, 2022; Lasser et al., 2023). A comparison between Twitter and German parliament data has also been made. Schaefer, Abels, Lewandowsky, and Stede (2023) used topic modelling to analyse how different parties communicated differently on Twitter on the topic of climate with the result that climate change communication does not differ substantially between Twitter and parliamentary speeches, but across the political spectrum. However, this study only came to our attention after we had conducted our analyses. Hence, we did not consider their methodological choices when making our own. Nonetheless, we found that their approach was very similar to ours.

[1] We will refer to the micro-blogging platform as Twitter, because the dataset used was collected before it was renamed.

2 Methods

In the following analyses, we used python 3.9.13, with the packages `numpy`=1.12.4, `pandas`=1.5.1, `bertopic`=0.15.0, and `SciPy`=1.9.2. [2]

2.1 BERTopic

We used the topic modelling technique BERTopic as a method to detect topics in our datasets (Grootendorst, 2022). BERTopic does so by applying the Sentence-BERT framework to generate high-dimensional vector representations capturing the sentence's meaning and emotional undertones. The embedding used is the `paraphrase-multilingual-MiniLM-L12-v2` which is trained on a multilingual corpus including German and results in a 384-dimensional vector. Following this, a dimensionality reduction technique, named UMAP, is applied to the vectors to reduce their sizes. This improves the performance of the subsequent clustering. The clustering technique used is Hierarchical Density-Based Spatial Clustering (HDBSCAN). Its benefit is, in comparison to other techniques, that it detects outliers in the process, allowing for documents to remain unclassified. By not forcing all documents in a topic, the coherence of the topics is improved.(reference?) Finally, class-based TF-IDF is utilised to create interpretable topic descriptions and lets us identify the most relevant keywords in each cluster.

2.1.1 Finetuning

Identifying the best set of parameters presents itself as quite a challenge given that it is difficult to quantify what one tries to optimise. The approach we followed was to maximise topic coherence while still minimising outliers. This was done via human evaluation. To be able to compare the two topics of the two datasets, our goal was additionally to keep the parameters as similar as possible and only change the parameters which are dependent on the corpus size or the average length of the corpus entries. A detailed explanation of the modified parameters can be found in the appendix (see Appendix A).

2.2 Data Preprocessing

The plenary protocols of the German parliament are openly accessible in the XML format provided by the German parliament (Deutscher Bundestag, 2019). Due to the complexity of the data encoding, we used a parser available on GitHub to parse the data and add it into a dataframe (Armin Pournaki, 2020). The parliament only started publishing the data as XML in the 19th legislative period. This period ranges from 19 October 2017 until 24 September 2021. Thus we decided to only use the data from that period. This results in a total of $24,666$ speeches. After plotting the length of the speeches, we observed a bimodal distribution of the data. The first peak contained mostly answers following speeches and answers in the "Fragestunde", a format in the parliament where politicians of the governing parties have to respond to questions from the opposition. Removing speeches with fewer than $2,500$ characters, we end up with a total of $19,018$ speeches with a mean length of $M = 3,904.85$ ($SD = 2,232.44$). No further data-cleaning step was applied.

The Twitter dataset was gathered by Lasser et al. (2022) and contains $1,559,359$ tweets from German politicians from 1 January 2016 to 11 February 2023. To match and be able to compare the topics to the parliamentary speeches, we only used the tweets from the 19th election period. Moreover, we decided to only take the tweets from active MPs which reduced the size to $708,874$. Minor preprocessing was applied to the data, namely removing hashtags, hyperlinks, emojis, and encoding symbols such as `\n` (new line). Furthermore, we changed the @handles into the string `user` to maintain the sentence structure while nevertheless removing the @handle. Following these steps, all tweets with fewer than three words have been removed due to lack of content resulting in 666,565 tweets with a mean length of $M = 151.59$ ($SD = 83.82$). Please note that BERTopic relies on the sentence embeddings and keeping the original structure of the text is important for transformer models (Egger & Yu, 2022).

[2] All code can be found at https://github.com/akristing22/Master_project_SDS

2.3 Topic Matching

The topic model trained on the parliament data discovered 132 topics, classifying $9,889$ (52%) of the speeches. The model trained on the Twitter data identified 98 topics, classifying $288,373$ (58%) of the tweets.

In order to assign the topics discovered in the parliament and Twitter corpora to their matching counterparts in the other corpus, we utilised the cosine similarity between the embeddings of topic representations as a measure of similarity between topics. For each topic from the parliament model, we identified the closest three candidates as potential matches from the Twitter topics. We then individually reviewed the keywords representing the respective topics and manually assigned which candidate from the Twitter topics was the best match for the parliament topic, or alternatively decided none of the candidates was a match. We used a majority vote to decide on the best match in case of disagreement between raters. The procedure had high interrater reliability, with Fleiss' $\kappa = .76$. Due to the nature of the procedure, it was possible that more than one parliament topic was matched to the same Twitter topic. In these cases, all parliament topics as well as the Twitter topic were included in the matched topic.

In this fashion, we initially obtained 49 matched topics, however, on visual inspection of the data, we identified an outlier topic that had an exceedingly high number of tweets $(45,824)$ associated with it. Upon further inspection of the topic's keywords, we concluded that it did not constitute a coherent topic and was rather a very broad collection of phrases used to discuss politics in general. We therefore excluded this topic from further analysis, resulting in a total of 48 matched topics, matching 31% of parliament speeches with 22% of tweets.

2.4 Super Topics

In addition, we decided to assign the documents from our corpora to a number of broader, overarching categories in order to assist with interpretability and gain further insights into the data. We refer to these categories as *super topics*, as they contain several of the topics discovered by the model. The basis for these super topics were the areas of responsibility of the different ministries of the German government from the analysed election period, although these responsibilities were at times split into multiple super topics if deemed more appropriate, resulting in 19 super topics plus an 'other' category (see Appendix B). These super topics were assigned manually to all topics discovered by the Twitter model and the parliament model. Again, this annotation was at first performed individually, with a majority vote then deciding the final assignment. Interrater reliability was acceptable for both Twitter ($\kappa = .60$) and parliament topics ($\kappa = .65$). We were able to ascertain an appropriate super topic for 44% of parliament speeches and 25% of tweets.

3 Analysis

In order to obtain a measure for the similarity in attention that topics received in parliament and on Twitter by MPs, we calculated a number of rank correlations. We opted for rank correlations over Pearson's correlation coefficient because the document counts did not follow a normal distribution. This was evident through a visual check of the data and confirmed by normal distribution tests. We performed Shapiro-Wilk tests on the distributions of topic and super topic sizes in both datasets, all of which indicated a significant departure from normality. We calculated both Spearman's ρ as well as Kendall's τ, since both are commonly used and neither offers a clear advantage over the other. A more fine-grained analysis was conducted on the level of the individual matched topics, while a more coarse-grained analysis was conducted on the level of super topics.

3.1 Topic Correlation

Calculating Spearman's ρ over the number of tweets and number of parliament speeches associated with the 48 matched topics, we found a moderate to strong, positive correlation that was statistically highly significant ($\rho = .54$, $p < .001$). Calculating Kendall's τ in a similar fashion, we likewise obtained a moderate to

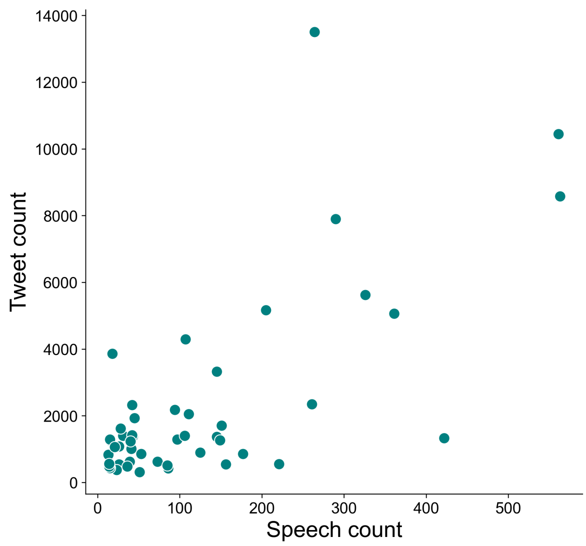

Figure 1: Matched topics with associated document counts for both corpora

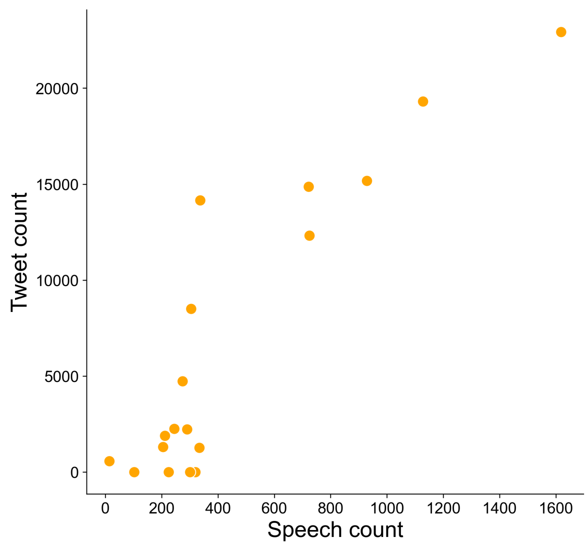

Figure 2: Super topics with associated document counts for both corpora

strong, positive correlation that was again statistically highly significant ($\tau = .38$, $p < .001$). The counts of parliament speeches and tweets associated with the same topics are plotted in Figure 1.

3.2 Super Topic Correlation

On the level of super topics, we calculated Spearman's ρ over the document counts in parliament and on Twitter relating to the respective 19 super topics. The correlation was found to be positive, and even stronger than on a topic level, while remaining highly statistically significant ($\rho = .74$, $p < .001$). Kendall's τ was also calculated on the super topic document counts, and again we found a strong, positive correlation between the number of tweets and the number of parliament speeches associated with the same super topics ($\tau = .62$, $p < .001$). The counts of parliament speeches and tweets associated with the same super topics are visualised in Figure 2.

4 Conclusion

The topic modelling gives insight into what has been discussed on Twitter and in parliament in the 19th election period by German MPs. Looking at the three top topics in each corpus already allows for some interpretation: The most discussed topics on Twitter were climate (4.7%), tax politics (3.6%) and Covid (3%). The three most discussed topics in parliament were Covid (5.7%), immigration (4.3%) and budget and finance (3.5%)[3]. Covid is a shared theme across both corpora, which makes sense considering the time frame of the data. It encompasses roughly the first two years of the Covid pandemic during which this topic dominated every conversation, everywhere. Interestingly, climate is the most discussed topic on Twitter but not in the top three topics in parliament. Conversely, immigration is one of the most discussed topics in parliament, but not on Twitter.

When looking at the super topics and the number of documents associated with them, depicted in Figure 3, it is apparent that the overall distributions are similar for both tweets and speeches. The parliament data is a little more spread out and covers 18 of the 19 super topics, while the tweets only cover 14 of them. It seems that, on Twitter, there is a bias towards discussing the bigger topics which might get more attention, while some topics which are a part of parliament discussions (albeit a minor), are not shared on Twitter at all (e.g.

[3]These are each the largest three coherent topics, leaving out topics which have no inherent political theme. The percentages denote the number of documents in each topic relative to the number of documents in all topics (without documents which have not been assigned any topic).

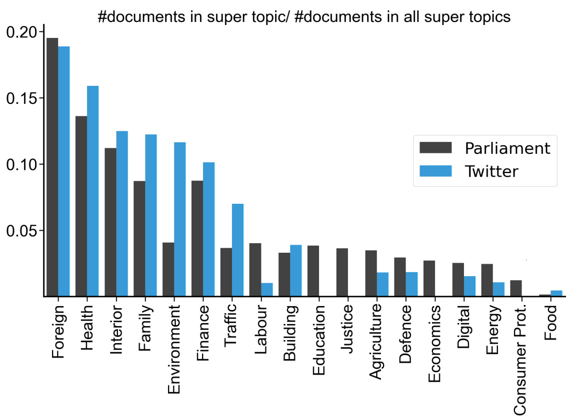

Figure 3: Distribution of super topic sizes

justice, education).

The biggest difference between Twitter and parliament regarding super topic distribution can be seen in the category environment. Climate and environment-related topics are discussed much more on Twitter than in parliament.

Coming back to our research question of whether the topics debated on Twitter by German politicians differ from the topics debated in the German parliament, we can conclude that they are overall similar as supported by the moderate to strong rank correlations between (super) topic distributions. However, there are some clear deviations as with the Environment topic. There seems to exist a small bias towards more 'high-profile' topics on Twitter, while the discussions in parliament are more inclusive of all varieties of topics.

5 Critique

The findings of the study project help us understand how German MPs make use of Twitter in terms of what topics they discuss there in contrast to what they actually discuss in parliament. However, there are some limitations to the findings which have to be considered. Additionally, some questions arose from the results which can be investigated in further research.

5.1 Limitations

An issue that we ran into very quickly is the vast difference of the two corpora that we analysed. Tweets are all less than 280 characters long, while our sample of parliament speeches starts at 2,500 characters. The register and grammar of the corpora are very different, as well as the size of the corpora themselves. BERTopic handles both types of documents relatively well, which helped mitigate some of the problems of having two very distinctive corpora. However, the difference in the nature of the corpora might have an impact on the quality and comparability of the topic modelling.

The topic modelling itself has some limitations. Almost half of the documents are not classified as any topic and after topic matching, we were able to analyse an even smaller sample. So when interpreting the results, one has to keep in mind that a significant share of documents is not being considered. Additionally, some of the topics found by BERTopic are not entirely coherent. For example, fishing industry and regulation are grouped together with shipbuilding (topic 100 of parliament data). While ships and fishing are certainly semantically related and thus their Sentence-BERT embeddings are close together, the actual contents are politically quite different.

There are also some drawbacks regarding the methods of topic matching and the matching of topics to super topics. For the matching between topics, we used a mixed-methods approach by combining cosine similarity to identify the three most similar Twitter topics for each parliament topic with human-in-the-loop evaluation of these best-fitting topics. With a Fleiss' $\kappa = 0.76$, the inter-rater reliability for choosing the best-fitting topic of the three closest topics is moderately high. Still, this approach is a little handwavy and relies heavily on our interpretation of the topics and how to define a 'match'.

The matching of topics to super topics is even more dependent on the interpretation of the (super) topics. By opting for the competencies of the ministries of the respective government, our approach is biased towards the parliament data. This might be a reason why 44% of the speeches could be classified into super topics, but only 22% of the tweets. One main issue with matching the topics to super topics is that they overlap in many aspects, as many topics have multiple underlying themes.

Take for example topic 93 of the parliament data with the representation *['abtreibungsgegnern', 'schwangerschaftsabbruch', 'abtreibungen',...]* (anti-abortionists, pregnancy termination, abortions,...). This topic has many dimensions, most prominently feminist (family affairs, senior citizens, women and youth), reproductive health (health) and legal (justice) issues. We unanimously decided for the first super topic (family affairs, senior citizens, women and youth), but you could argue for either one. We tried to be consistent with these kinds of decisions, but our classification is certainly not the only way to go about assigning the super topics.

Additionally to the heavy overlap between (super) topics, some topics are not under the purview of any federal ministry, either because they are the responsibility of the states or not governmentally regulated (e.g. journalism). Still, inter-rater reliability as measured with Fleiss' κ was moderately high: .60 for matching of tweet topics to super topics and .65 for speeches.

When interpreting the difference between topics between Twitter and parliament, be aware of the fact that we did not analyse whether each individual MP focuses on different topics on Twitter or in parliament, but which topics are discussed by all MPs, respectively. The latter might be influenced by the varying numbers of speeches/tweets per MP. Firstly, the distribution of the number of documents per MP is not uniformly distributed, but there are few MPs with many tweets/speeches and many MPs with few tweets/speeches. This power-law-like behaviour is stronger for Twitter than for parliament. Secondly, the distribution of documents per MP only moderately correlates between the two corpora (Spearman's $\rho = .25, p < .001$). The MPs who tweet a lot are not necessarily the ones who hold many speeches in parliament. When trying to explain the differences between Twitter and parliament, one should pay attention to these differing distributions.

5.2 Future Research

The results of this analysis give rise to further possible investigations of MPs' behaviour on Twitter and in parliament. One line of research could be how politicians talk about certain topics depending on their party affiliations, as did Schaefer et al. with climate change related topics in their 2023 study. Additionally, comparing different legislation periods and how topics change over time would be possible. Specifically, it would be interesting to see if the discrepancy between Twitter and parliament regarding climate topics holds in the current (20th) legislation period.

Methodologically, it would be possible to allow fuzzy matching of topics to super topics, so to match a topic with multiple super topics, if sensible. Then, a more accurate analysis of the topics might be possible, as well as an analysis of which super topics are more related than others.

Another intriguing line of research would be to investigate the temporal relationship of topics on Twitter and in parliament and whether one influences the other. Because of the difference in time-resolution (24/7 vs 2-3 sessions per week), this is a challenging question to tackle. However, it would allow for a better understanding of the influence social media has on parliamentary proceedings.

In order to understand why some topics are more heavily discussed than others on Twitter, one could analyse the popularity of the tweets depending on their (super) topic. Possibly, MPs opt for tweeting more about topics which engage more people and thus leave out some other topics which might not get as much attention.

References

Armin Pournaki. (2020). *bundestag-parser*. `https://github.com/pournaki/bundestag-parser`. (Accessed: 04.11.2023)

Benoit, K., Schwarz, D., & Traber, D. (2012). The sincerity of political speech in parliamentary systems: A comparison of ideal points scaling using legislative speech and votes. In *2nd annual conference of epsa, berlin* (pp. 19–21).

Bright, J., Hale, S., Ganesh, B., Bulovsky, A., Margetts, H., & Howard, P. (2020). Does campaigning on social media make a difference? evidence from candidate use of twitter during the 2015 and 2017 uk elections. *Communication Research*, *47*(7), 988–1009.

Contreras, K., Verbel, G., Sanchez, J., & Sanchez-Galan, J. E. (2022). Using topic modelling for analyzing panamanian parliamentary proceedings with neural and statistical methods. In *2022 ieee 40th central america and panama convention (concapan)* (pp. 1–6).

Curran, B., Higham, K., Ortiz, E., & Vasques Filho, D. (2018). Look who's talking: Two-mode networks as representations of a topic model of new zealand parliamentary speeches. *PloS one*, *13*(6), e0199072.

Deutscher Bundestag. (2019). *Endgültige plenarprotokolle*. `https://www.bundestag.de/dokumente/protokolle/plenarprotokolle`. (Accessed: 02.11.2023)

Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in sociology*, *7*, 886498.

Frid-Nielsen, S. S. (2018). Human rights or security? positions on asylum in european parliament speeches. *European union politics*, *19*(2), 344–362.

Garimella, K., & Weber, I. (2017). long-term analysis of polarization on twitter in: Proceedings of the 11th international conference on web and social media (icwsm); may 15 2017-may 18. *Montreal, Canada*.

Geese, L. (2020). Immigration-related speechmaking in a party-constrained parliament: Evidence from the 'refugee crisis' of the 18th german bundestag (2013–2017). *German Politics*, *29*(2), 201–222.

Golbeck, J., Grimes, J. M., & Rogers, A. (2010). Twitter use by the us congress. *Journal of the American society for information science and technology*, *61*(8), 1612–1621.

Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Kreiss, D. (2016). Seizing the moment: The presidential campaigns' use of twitter during the 2012 electoral cycle. *New media & society*, *18*(8), 1473–1490.

Kruikemeier, S. (2014). How political candidates use twitter and the impact on votes. *Computers in human behavior*, *34*, 131–139.

Lasser, J., Aroyehun, S. T., Carrella, F., Simchon, A., Garcia, D., & Lewandowsky, S. (2023). From alternative conceptions of honesty to alternative facts in communications by us politicians. *Nature human behaviour*, *7*(12), 2140–2151.

Lasser, J., Aroyehun, S. T., Simchon, A., Carrella, F., Garcia, D., & Lewandowsky, S. (2022). Social media sharing of low-quality news sources by political elites. *PNAS nexus*, *1*(4), pgac186.

Marcinkiewicz, K., & Stegmaier, M. (2019). Speaking up to stay in parliament: the electoral importance of speeches and other parliamentary activities. *The Journal of Legislative Studies*, *25*(4), 576–596.

Schaefer, R., Abels, C., Lewandowsky, S., & Stede, M. (2023). Communicating climate change: a comparison between tweets and speeches by german members of parliament. In *Proceedings of the 13th workshop on computational approaches to subjectivity, sentiment, & social media analysis* (pp. 479–496).

Schürmann, L., & Stier, S. (2023). Who represents the constituency? online political communication by members of parliament in the german mixed-member elec-

toral system. *Legislative Studies Quarterly*, *48*(1), 219–234.
Silva, B. C., & Proksch, S.-O. (2022). Politicians unleashed? political communication on twitter and in parliament in western europe. *Political science research and methods*, *10*(4), 776–792.
Stier, S., Bleier, A., Lietz, H., & Strohmaier, M. (2020). Election campaigning on social media: Politicians, audiences, and the mediation of political communication on facebook and twitter. In *Studying politics across media* (pp. 50–74). Routledge.
Udanor, C., Aneke, S., & Ogbuokiri, B. O. (2016). Determining social media impact on the politics of developing countries using social network analytics. *Program*, *50*(4), 481–507.
Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., & Li, X. (2011). Comparing twitter and traditional media using topic models. In *Advances in information retrieval: 33rd european conference on ir research, ecir 2011, dublin, ireland, april 18-21, 2011. proceedings 33* (pp. 338–349).

Appendix

A Hyperparameter

Model	Parameter name	Tweets	Debate	Explanation
UMAP	n_components	2	60	It is recommended to scale the parameter based on the complexity of the content. Since the average debate was 30x longer than the average length of a tweet we set the value 30x higher.
UMAP	metric	cosine	cosine	Due to the high dimensionality of the data cosine is recommended.
UMAP	min_dist	0.2	0.2	This parameter controls level of clumpiness and granularity in the data.
UMAP	n_neighbours	15	15	This parameter controls local versus global structures in the data.
TF-IDF	bm25_weighting	True	True	Lead to more robust results regarding the stopwords, especially in smaller datasets.
TF-IDF	reduce_frequent_words	True	True	Helps to reduce the impact of words that appear too frequently.
BERTopic	representation_model	KeyBERT	KeyBERT	Use a KeyBERT-like model to fine-tune the topic representations to extract keywords of the topics
BERTopics	min_topic_size	300	12	Scaled according to the dataset size.

B Super Topics

The following 19 super topics were used to further categorise the topics:

- Agriculture
- Building
- Consumer Protection
- Defence
- Digital Infrastructure
- Economics
- Economic Cooperation and Development
- Education and Research
- Energy
- Environment, Nature Conservation and Nuclear Safety
- Family Affairs, Senior Citizens, Women and Youth
- Finance
- Food
- Foreign Affairs
- Health
- Interior and Community
- Justice
- Labour and Social Affairs
- Transport