

Exploring Two Decades of Parliamentary Speeches on the Use of Narratives

Matti Wiegmann,¹ Jürgen Neyer,² Benno Stein³

¹Kassel University, ²European University Viadrina, ³Bauhaus-Universität Weimar
firstname.lastname@{uni-kassel.de | europa-uni.de | uni-weimar.de}

Abstract

Political scientists are interested in changes in political discourse over time. However, the topics of interest, such as the changes in support for or understanding of certain narratives, are often ill-defined and require deliberation, which prevents most lexical or metadata-based methods of temporal aggregation. To enable a diachronic analysis, we propose to model such settings as a series of binary document classification tasks – which current reasoning LLMs can adequately solve – and aggregate the decisions into a temporal signal. Specifically, we propose to use LLMs to classify if a parliament speech is in support of either of two narratives, and we use the monthly count of positives per narrative to track the support over time. We show that the classification is sufficiently accurate and use it to create detailed time series data showing support for the selected narratives in speeches given in the European Parliament from 2006 to 2023. The method is developed in close collaboration with political scientists and is considered an ideal starting point for diachronic analyses of political decision-making processes by domain experts.

Keywords: narratives, diachronic analysis, NLP for political sciences

1. Introduction

Narratives in political science are stories that, through repetition and variation, are “embraced by a group that also tells, in one way or another, something about that group” (Shenhav, 2015). In this sense, the narratives of groups, parties, and nations express the identity of a group and the beliefs of its members regarding sovereignty, i.e., who holds power and who should hold power.

Narratives are typically studied via extensive, qualitative analysis of political speech, either to formulate a narrative or to study its development or application in political discourse. An in-depth narrative analysis lays out the underlying elements, the series of events, the agents and their motivation, mode and process of communicating, repeating, and varying the narratives, and what the totality of these elements says about the group that embraces it (Shenhav, 2015). Consider the example of Bueger (2013), who analyzes interviews with Somali pirates to extract the narrative of piracy “as a quasi-state practice of the protection of sovereignty against foreign intruders” as a justification of who should have power in the region.

Such qualitative analyses become impossible when narratives are very extensive and too many documents need to be analyzed, either because the narrative is so fundamental that it influences much of the political discourse, or because the narrative is relevant over a long period of time. Both restrictions apply to narratives of sovereignty.

However, technology-assisted, quantitative methods can reliably answer certain research questions about narratives—especially when a statements about a trend are more important than high

(classification) accuracy at the per-document level. This is often the case with diachronic questions: How does the distribution (or, in the much more difficult case, the meaning) of a narrative vary over time or in relation to specific events? One example of this is the study by Bleakley (2021) on the development of the “Pizzagate” conspiracy narrative, which is based on counting tweets in three predefined topic categories over several years.

In this paper, we propose a technology-assisted approach to analyze the prevalence of two narratives on European sovereignty (De Vries, 2023):

- | |
|--|
| (1) Supranationalism ~ EU institutions should have power. |
| (2) Nationalism ~ Nation states should have power. |

Specifically, we use large language models (LLMs) to do a distant reading of the parliamentary debates to provide a prevalence signal for a diachronic analysis. We then develop a time series visualization to help political scientists develop data-driven hypotheses and assess in which cases our method delivers interpretable and meaningful analysis results and in which cases it fails.

We contribute a prompt-based classifier to detect traces of a narrative in a parliament speech and show its efficacy on a small, expert annotated dataset (Section 3). The classifier scores a balanced accuracy of about 0.85 for large models (Qwen-32B) and 0.75 for more efficient models (Qwen-8B), which we argue is sufficiently accurate for a prevalence analysis. We conduct a diachronic analysis of the prevalence of the two aforementioned narratives across a corpus of 137 844 speeches given in the European Parliament (EP) between 2006 and 2023. The corpus

is split into topics using BERTopic, the speeches in the most common topic clusters are classified, and the resulting prevalence ratio approximated across monthly samples. Our visualization shows the results in absolute and relative relations and in context with the significant events (Section 4).

We find that the resulting time series show meaningful differences across time in prevalence (How common the narratives are in the speeches?), density (How frequently is the topic discussed?), and ratio (Which narrative is more common?), and that these correspond to related world events. We also point out the limitations of our method: certain topics seem to lean toward one of the two themes, and a high density is required to obtain meaningful results (Section 4.3).

2. Related Work

Narratives are a frequently studied topic in literary and social sciences with many facets, from the structuralist analysis of stories (Genette, 1980; Abbott, 2008) to complex and often tacit “social narratives” (Shenhav, 2015) of nations or “meta-narrative” worldviews (Lyotard, 1994; Kaplan et al., 2022). The narratives we study here belong to the latter category, as concrete events, actors, and goals are often only implicitly known to the speaker, but the narrative is recognizable to the audience.

Computational work on narrative analysis focuses on narrative understanding (Piper et al., 2021), such as detecting a narrative structure or individual elements: Graaf et al. (2016)’s analysis of health-related stories in advertisements, Falk and Lapesa (2023)’s collection of narratives used for argumentation. Many related works also focus on conspiracy narratives, and, for example, parse structural elements (actors, goals, victims, etc.) of conspiracy narratives Korenčić et al. (2024), extract actors and their relationships from social media for the Bridgegate and Pizzagate conspiracies (Tangherlini et al., 2020), or extract a graph of events and their relationships from long-form news and use a graph-based classifier to detect conspiracy narratives (Lei and Huang, 2023).

However, we are more interested in recognizing (support for) a predefined narrative across many texts than in extracting (elements of) a single narrative from a single text. In that direction, Zhao et al. (2023) presents a temporal analysis by analyzing changes of COVID-19 narratives on social media over time, but focusing on elements in the narrative change instead of pure prevalence. Similarly, Moffitt et al. (2021) detect social media posts supporting a pre-defined COVID narrative and aggregate the results, but across communities instead of time (Moffitt et al., 2021).

3. Narrative Classification

Our basis for a diachronic analysis is an LLM classifier for document-level narrative detection. The task is formulated as a one-class reasoning task: Given a parliamentary speech and the definition of a narrative, decide whether the narrative is supported. This is a complex problem, as the “support relation” between a narrative and a speech can be multifaceted, e.g. by referring positively to actors or goals, by presenting a current issue in line with the narrative, or by affirming shared values. In particular, narratives are rarely explicitly named. However, the task is solvable, as expert annotators reach a reasonable agreement. Figure 3 shows some examples with detailed explanations.

3.1. Data

The corpus used in this work consists of 137,844 speeches delivered in the European Parliament between 2006 and 2022, collected by Husmann and Wiegmann (2025). The speeches are split into 474,545 paragraphs to isolate individual political statements and facilitate unique topic clustering.

	Supranationalism	Nationalism
Positive	152	55
Negative	30	130
Disagreement	18 ($\alpha = 0.67$)	15 ($\alpha = 0.79$)
Σ Annotated	200	200

Table 1: Descriptive statistics of the test dataset.

Test Data To evaluate the efficacy of the narrative classification, we create a small, manually annotated test dataset, shown in Table 1. Altogether, the dataset contains 400 examples, annotated as positive or negative for one of the two narratives.

The examples are selected via keyword-filtering: For each narrative, we randomly draw 200 paragraphs from the corpus that contain at least one of the actors, events, and purposes contained in the narrative definition.

The test data includes the 367 examples that are unanimously annotated as positive or negative by two trained graduate students of political science, examples with disagreement are discarded. The annotator agreement is high with 0.67–0.79 Krippendorff’s α . The annotation guidelines were created by the annotators and the consulting political scientist on hold-out examples. The students have been employed for the duration of the campaign and standard rates for student assistants.

3.2. Prompts

The prompt, shown in Figure 4 (Appendix A), consists of five parts: an initial task statement, the narrative definition, a classification guideline, and in-context learning (ICL) examples.

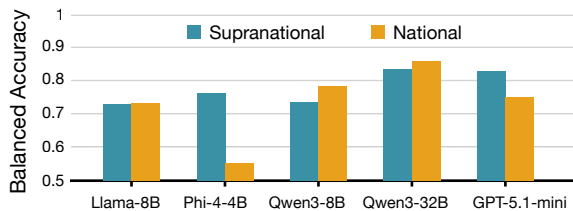


Figure 1: Model comparison for both narratives.

The definitions, shown in Figure 5 (Appendix A), were derived by expert political scientists based on relevant work in this field: (Sweet and Sandholtz, 1997; Rittberger, 2003; Mancini, 1998) for supranational, (Moravcsik, 2013; Tallberg, 2008; Pollack, 2003) for national. The classification guideline was derived from the annotation guide and consists of two filter instructions prompting the removal of inappropriate or off-topic examples, a Chain-of-thought statement prompting the creation of an explanation, and an instruction prompting a ‘yes’ or ‘no’ answer. The four included ICL examples, two positive and two negative, were manually selected from the source corpus; the reasoning for each was manually written.

3.3. Evaluation

We evaluate the performance of five current LLMs on the test dataset: Llama-8B, Phi-4-4B, Qwen-3 8B and 32B, and GPT-5.1-mini. This selection includes generally well-performing open-weight reasoning models of different sizes, as well as GPT-5.1 as a reasonable SotA comparison and Llama as a non-reasoning baseline.

The effectiveness is evaluated via balanced accuracy (BA), as the dataset is small and imbalanced and so the F_1 score is distorted by small changes to minority class assignments.

Results The most effective model, as shown in Figure 1, is Qwen-3-32B with a balanced accuracy of around 0.85. In general, the larger models perform better, as expected, but are also more expensive: Qwen in terms of runtime and GPT in terms of API costs. Since our goal is to use the model for diachronic analysis, which requires classifying a large number of examples, we chose Qwen3-8B because it offers the best compromise between effectiveness and efficiency.

4. Diachronic Analysis

We collect the required time series data for a diachronic analysis of narratives by aggregating the document-level decisions for the entire 17-year period covered by the source corpus by month. In this regard, we aggregate speeches per-topic for the 15 most common topics. The Figures 2 and 6 visualize the results.

4.1. Topic Modeling

BERTopic (Grootendorst, 2022) is used to split the corpus into individual topics; all further processing and analyses is done individually for selected topics, where the rationale is as follows. Firstly, the corpus is very large, with about 470,000 paragraphs, and classifying each speech for each narrative is too expensive. Secondly, narrative support varies between topics (i.e., a group may favor nationalism on one topic and supranationalism on another). Thirdly, there are several topics that are not relevant to narrative analysis and would obscure the actual signal (e.g., rhetorical elements such as greetings and acknowledgements, or rapporteur reports).

We apply BERTopic with the `all-mpnet-base-v2` embedding model and the default components: UMAP dimensionality reduction to 10 components considering 20 neighbors, HDBSCAN clustering with a minimum 100 documents, vectorization considering 1-to-3 grams, minimum document frequency of 20%, and a maximum document frequency of 60%. The noted hyperparameters were iteratively refined until the clusters, as indicated by the c-TFIDF and KeyBERT representations, were consistent, not too broad, but not more specific than individual debates. The smallest topics are, for example, “Rights of Polish women” and “Fast privacy and data protection”, while the largest topics are shown in the figures.

4.2. Temporal Aggregation

The classification results are aggregated per month as absolute values and as a ratio between the narratives. The absolute values are the number of passages classified as positive for the supranational narrative, the national narrative, and neither. The ratio is the number of supranational positive cases relative to all positive cases, with only months with at least 10 positive cases in total being counted.

Three additional key figures are calculated for interpretation purposes: smoothed ratio, density, and prevalence. The *smoothed ratio* is a B-spline approximation with weights close to zero for months with fewer than 10 positive values in total, resulting in a linear approximation over periods with sparse data. Smoothing helps to identify trends in the visualization, as monthly ratios tend to be noisy and vary greatly in less dense periods. *Density* is the relative number of months with at least 10 positives and indicates whether a topic is discussed continuously, i.e., how dense the temporal coverage of the data is. *Prevalence* is the ratio of positive passages to all passages over the entire period and indicates whether the narratives play a role within the topic.

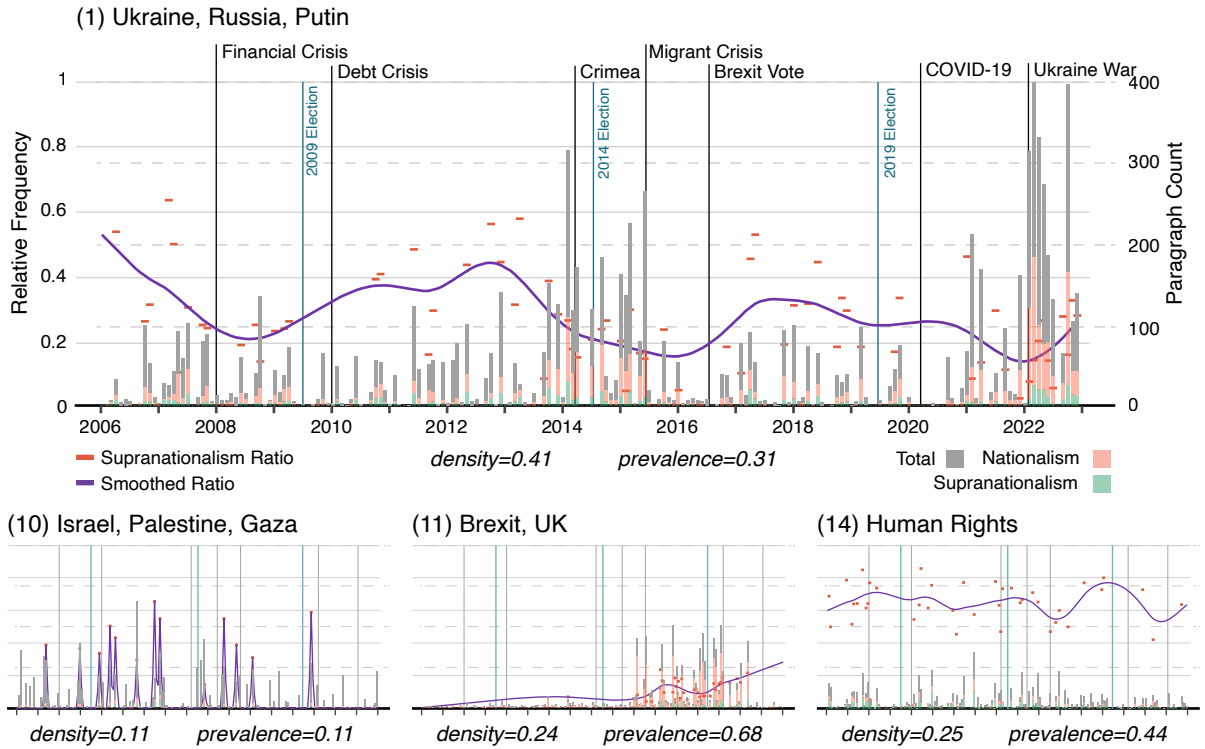


Figure 2: Visualization of the temporal aggregation for the most common topic as an example for the general case. The topics 10, 11, and 14 are shown as examples for unexpected results or biases. ρ is the Spearman correlation, d is the density, and p is prevalence.

4.3. Results and Discussion

The resulting time series data, both absolute and relative, are visualized in Figures 2 and 6 (Appendix B), with additional markers indicating elections and important events that influence European politics. Please note that these events are not exhaustive, as important legislative procedures or consultation phases are not listed.

Several observations can be made regarding the suitability and limitations of our method.

Firstly, there are particularly intense periods when a topic is discussed frequently, i.e. when a topic becomes urgent due to a crisis or new legislation, and less intense periods when topics are discussed less frequently and depend on specific debates. In addition, there are many gaps due to parliamentary breaks. As a result, the narrative ratio is unstable and varies greatly during sparse periods, and the smoothed ratio is not very robust, which can be seen for all of the topics we have presented. A clear example of this is topic (10). Considering the density is essential to avoid drawing false conclusions from noisy data.

Secondly, the prevalence of narratives in discourse depends on the topic and plays a much greater role for topics such as (7) Migration and (11) Brexit than in topics such as (4) Climate Change and foreign policy (1, 5, 10).

Thirdly, some topics are strongly influenced by one narrative or the other, such as (11) Brexit and

(14) Human Rights. While this is to be expected, given that Brexit is a typically nationalist idea and human rights are a core argument of supranationalism, it may also indicate a bias caused by the clustering.

Finally, the discourse is probably less nationalist than suggested in the plots. Our method probably underestimates the frequency of supranational narratives and overestimates national narratives, as the classifier tends to predict more false negative supranational examples, where positive cases predominate in the test data, while predicting more false positive national examples, where negative cases predominate.

5. Conclusion

This paper presents an approach for examining the fluctuating support for certain narratives in European Parliament debates over the past two decades. To this end, LLMs are used to identify support in individual speeches, with signals aggregated over a specific period of time. We develop a sufficiently efficient and effective prompt for Qwen3-8B and present derived time series data for several important debate topics, enabling political scientists to discover new hypotheses. We find that narratives should be examined separately by topic, that there are many periods in which data may be unreliable, and that even basic narratives are not equally prevalent for all topics.

6. Bibliographical References

- H. Porter Abbott. 2008. *The Cambridge Introduction to Narrative*. Cambridge University Press.
- Paul Bleakley. 2021. [Panic, pizza and mainstreaming the alt-right: A social media analysis of pizzagate and the rise of the qanon conspiracy](#). *Current Sociology*, 71:509 – 525.
- Christian Bueger. 2013. [Practice, pirates and coast guards: the grand narrative of somali piracy](#). *Third World Quarterly*, 34(10):1811–1827.
- Catherine E. De Vries. 2023. How foundational narratives shape european union politics. *JCMS: Journal of Common Market Studies*, 61(4):867–881.
- Neele Falk and Gabriella Lapesa. 2023. Storyarg: a corpus of narratives and personal experiences in argumentative texts. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2350–2372.
- G rard Genette. 1980. *Narrative discourse: An essay in method*, volume 3. Cornell University Press.
- Anneke de Graaf, Jos  Sanders, and Hans Hoeken. 2016. Characteristics of narrative interventions and health effects: A review of the content, form, and context of narratives in health-related narrative persuasion research. *Review of communication research*, 4:88–131.
- Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.
- Sven Husmann and Matti Wiegmann. 2025. [European parliament speeches 2002-2022](#).
- Yael R Kaplan, Tamir Sheafer, and Shaul R Shenhav. 2022. Do we have something in common? understanding national identities through a meta-narrative analysis. *Nations and Nationalism*, 28(4):1152–1172.
- Damir Koren ci , Berta Chulvi, Xavier Bonet Casals, Alejandro Toselli, Mariona Taul , and Paolo Rosso. 2024. What distinguishes conspiracy from critical narratives? a computational analysis of oppositional discourse. *Expert Systems*, 41(11):e13671.
- Yuanyuan Lei and Ruihong Huang. 2023. [Identifying conspiracy theories news based on event relation graph](#). *ArXiv*, abs/2310.18545.
- Jean-Fran ois Lyotard. 1994. The postmodern condition. *The postmodern turn: new perspectives on modern theory*, pages 27–38.
- G Federico Mancini. 1998. Europe: the case for statehood. *Eur. LJ*, 4:29.
- J. D. Moffitt, Catherine King, and Kathleen M. Carley. 2021. [Hunting conspiracy theories during the covid-19 pandemic](#). *Social Media + Society*, 7.
- Andrew Moravcsik. 2013. *The choice for Europe: Social purpose and state power from Messina to Maastricht*. Routledge.
- Andrew Piper, Richard Jean So, and David Bamman. 2021. [Narrative theory for computational narrative understanding](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 298–311. Association for Computational Linguistics.
- Mark A. Pollack. 2003. *The engines of European integration: Delegation, agency, and agenda setting in the EU*. OUP Oxford.
- Berthold Rittberger. 2003. The creation and empowerment of the european parliament. *JCMS: Journal of Common Market Studies*, 41(2):203–225.
- Shaul Shenhav. 2015. *Analyzing social narratives*. Routledge.
- Alec Stone Sweet and Wayne Sandholtz. 1997. European integration and supranational governance. *Journal of European public policy*, 4(3):297–317.
- Jonas Tallberg. 2008. Bargaining power in the european council. *JCMS: journal of common market studies*, 46(3):685–708.
- Timothy R. Tangherlini, Shadi Shahsavari, Behnam Shahbazi, Ehsan Ebrahimzadeh, and Vwani P. Roychowdhury. 2020. [An automated pipeline for the discovery of conspiracy and conspiracy theory narrative frameworks: Bridgegate, pizzagate and storytelling on the web](#). *PLoS ONE*, 15.
- Wanying Zhao, Fiona Guo, Kristina Lerman, and Yong-Yeol Ahn. 2023. [Discovering collective narratives shifts in online discussions](#). In *International Conference on Web and Social Media*.

A. Prompts

Positive instance: The Commission did not hesitate to act in its role of guardian of the treaties against a national legislation that restricted the right to freedom of association of civil society organisations by limiting their access to funding, in breach of EU law and of the Charter, as confirmed in 2020 by the Court of Justice.

Reasoning: The paragraph clearly belongs to the supranational narrative. The speaker refers to two of the three main actors in the supranational narrative: the EU Commission and the European Court of Justice. The EU Commission acts in its role as guardian of the treaties and monitors compliance with them, supported by the European Court of Justice, which has issued a judgment in this legal dispute.

Negative instance: We had a rule of law that had evolved since the days of Alfred The Great, but English law is now superseded by EU law. Impartial English courts are now overridden by the political courts of the European Court of Justice and the European Court of Human Rights. Our ancient protections of individual liberties such as Habeas corpus have been swept away by such measures as the European Arrest Warrant.

Reasoning: The paragraph does not correspond to the supranational narrative. The speaker refers to one of the three main actors in the supranational narrative: the European Court of Justice. The speaker questions the primacy of EU law over national, British law and sees the European Court of Justice as a threat to the rule of law and individual freedoms.

Figure 3: A positive and a negative example for the **supranational narrative** with reasoning provided by expert political scientists.

You classify political narratives in speeches. For a given passage from a political speech, decide if the speech supports the **[narrative]** narrative.

****Narrative Definition:****

[definition]

****Classification Task:****

Steps for analyzing and classifying speech passages:

1. Decide if the passage is unsubstantial, such as greetings, interludes, or unclear statements. In this case, answer with 'no'.
2. Decide if the passage has nothing to do with the narrative or its actors. In this case, answer with 'no'.
3. Draft an explanation for if the passage supports the narrative or not, citing the passage content where appropriate.
4. Give your answer as 'yes' or 'no'.

Consider the following examples:

****Passage:****

[ICL text]

****Explanation:****

[ICL reasoning]

The answer is **[ICL truth]**

Now analyze and classify this speech. Remember to finish your answer by stating the class as 'yes' or 'no'.

****Passage:****

[passage text]

Figure 4: The prompt template used to classify the speech passages. Bold text is variable.

National: The national narrative champions the nation-states and the European Council as the leading European actors. Member states are portrayed as the leading political institutions and the drivers of European politics. The national narrative favors cooperation and negotiation among the member states and these member states control the supranational institutions. Whilst cooperation and negotiation among the governments is deemed crucial to the legitimate functioning of the EU, the control of the supranational institutions is an essential purpose. This narrative is underpinned by the assumption that preserving national sovereignty and borders is critical to maintaining cultural diversity and national identity.

Supranational: The supranational narrative employs the supranational institutions as the main actors: the European Commission, the European Parliament, and the European Court of Justice. The European Parliament is the strongest protagonist of supranational democracy that aims for ever closer integration of states in the EU. Since it is directly elected, it guarantees that EU policies are democratically legitimated. The Commission acts as the 'Guardian of the Treaties', making fair rules for all and identifying violations. The European Court of Justice settles legal disputes and imposes fines in case of ongoing violations of the legal order of the EU. In the supranational narrative, these institutions collaborate to deepen market integration and strengthen the rule of law, democracy, and human rights.

Figure 5: Narrative definitions used in the prompt templates.

B. Additional Results

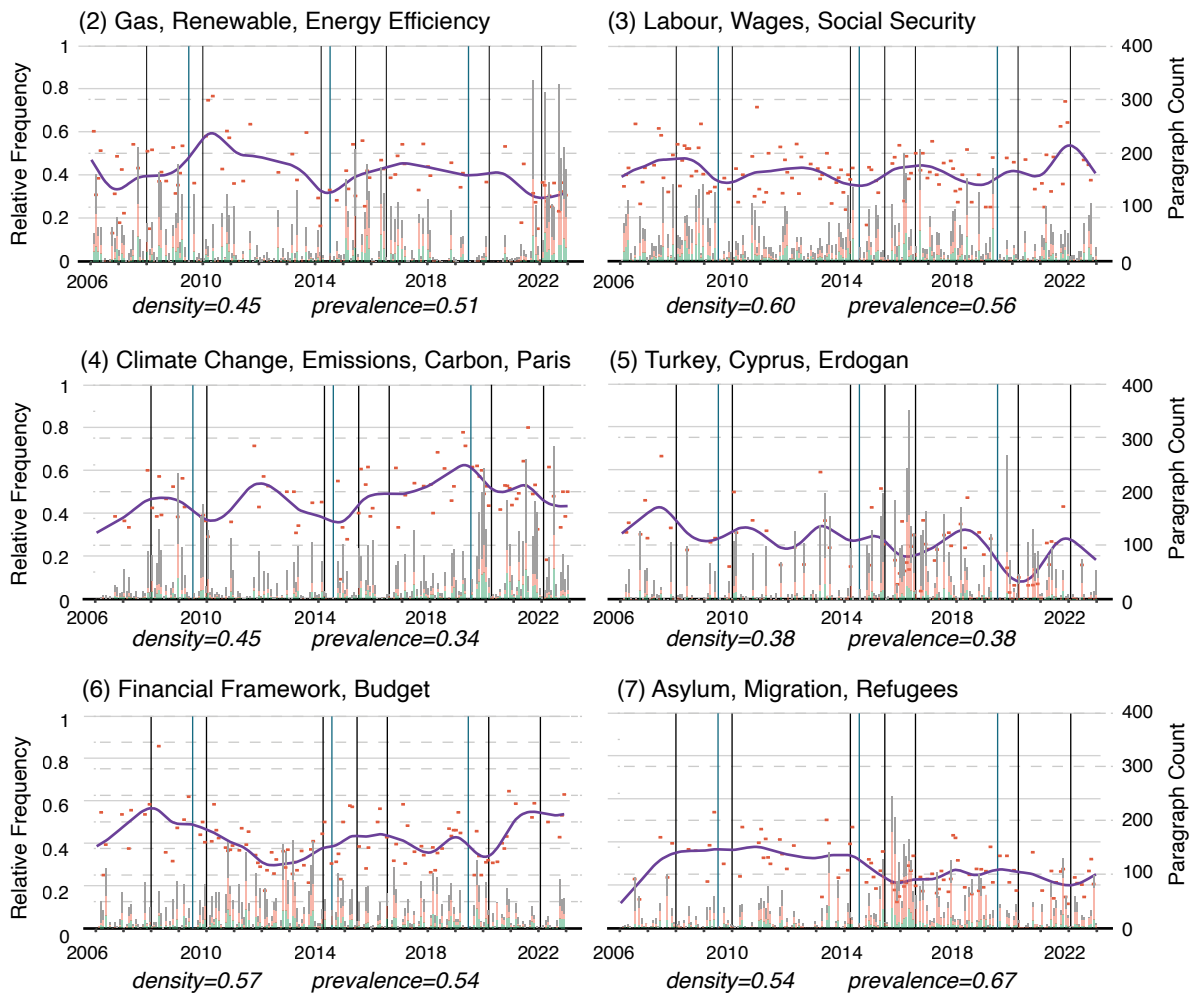


Figure 6: Result plots of the temporal aggregation for the topics 2–7.