Splitting Lips

Polarization through Parliamentary Speech

DIGITAL HUMANITIES HACKATHON HELSINKI

Anna Kryvenko, Ari-Heikki Rintaniemi, Bojan Evkoski, David Bordon, Ekaterina Glazacheva, Jani Marjanen, Katja Meden, Mark Mets, Nikoletta Jablonczay, Paul Pope, Pontus Hedlund, Risto Turunen, Topi Ranta, Vadym Kuzyak, Vid Klopčič

Polarization

Definitions of polarization vary (see Palonen 2009; Goet, 2019) and are commonly referred to:

- **a state of political division in society**
- **the act of creating political division through discourse**

We focus on **polarization as a state of affairs in parliament**. While individual methods cannot grasp the whole phenomenon, they can illuminate various aspects.

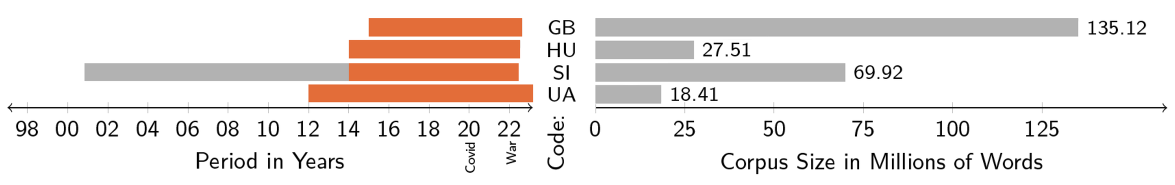
Manifestations	Approaches
groups do not talk about the same topics	sentence embeddings, TF-IDF
groups valuate the same topics differently	sentiment analysis
groups use alternative terms to set the agenda	close reading

Research questions

- How do specific topics polarize parliaments?
- How does polarization manifest over time?
- How can polarization be measured with computational methods?

Data

The research is based on the data set **Multilingual comparable corpora of parliamentary debates ParlaMint 3.0** (beta version) pre-released by CLARIN ERIC. We focused on **Great Britain, Hungary, Ukraine,** and **Slovenia** in the periods highlighted below.



Methods

- Extract three thematic subcorpora using LDA and corpus keyword methods: **European Union, War,** and **Healthcare.**
- Calculate speech representations and sentiment with multilingual **LLMs.**
- **Calculate and visualize SBERT embeddings** for finding thematic differences between political parties.
- Conduct **in-depth reading** of relevant speeches and refine information from the visualizations into interpretations.

Theme	Topic	Focus %	Sentiment
		Great Britain	
Theme	**Topic**	**Focus % (CON / LAB)**	**Sentiment (CON / LAB)**
EU	Brexit Referendum	11.1 / 17.6	0.10 / -0.08
War	Ukraine-Russia War	13.4 / 10.6	-0.08 / -0.30
Healthcare	Covid	30.1 / 22.5	0.17 / -0.35
		Ukraine	
		Focus % (Pro-UA* / Pro-RU*)	**Sentiment (Pro-UA* / Pro-RU*)**
EU	Language Policy	10.30 / 47.50	-0.34 / -0.55
War	Legislations in War	9.60 / 22.40	-0.41 / -0.58
Healthcare	Organ Transplantation	16.80 / 1.10	0.03 / -0.88
		Hungary	
		Focus % (Fidesz-KDNP / MSZP)	**Sentiment (Fidesz-KDNP / MSZP)**
EU	Corruption Charges	16.70 / 30.9	-0.38 / -0.48
War	Constitution Defense	13.4 / 29.5	-0.32 / -0.52
Healthcare	Covid	24.3 / 9.5	-0.13 / -0.52
		Slovenia	
		Focus % (SDS / SD)	**Sentiment (SDS / SD)**
EU	Tax Coffers	6.90 / 10.8	-0.04 / 0.02
War	Veteran Pensions	2.2 / 5.0	-0.30 / -0.32
Healthcare	Healthcare System	14.2 / 8.0	-0.26 / -0.48

Pro-Ukraine* - фБПП.фЄС.фУДАР
Pro-Russia* - фОПЗЖ.фОпоблок.фПР

Selected BERTopic subtopics highlighting the contrasting focuses between opposing political groups.

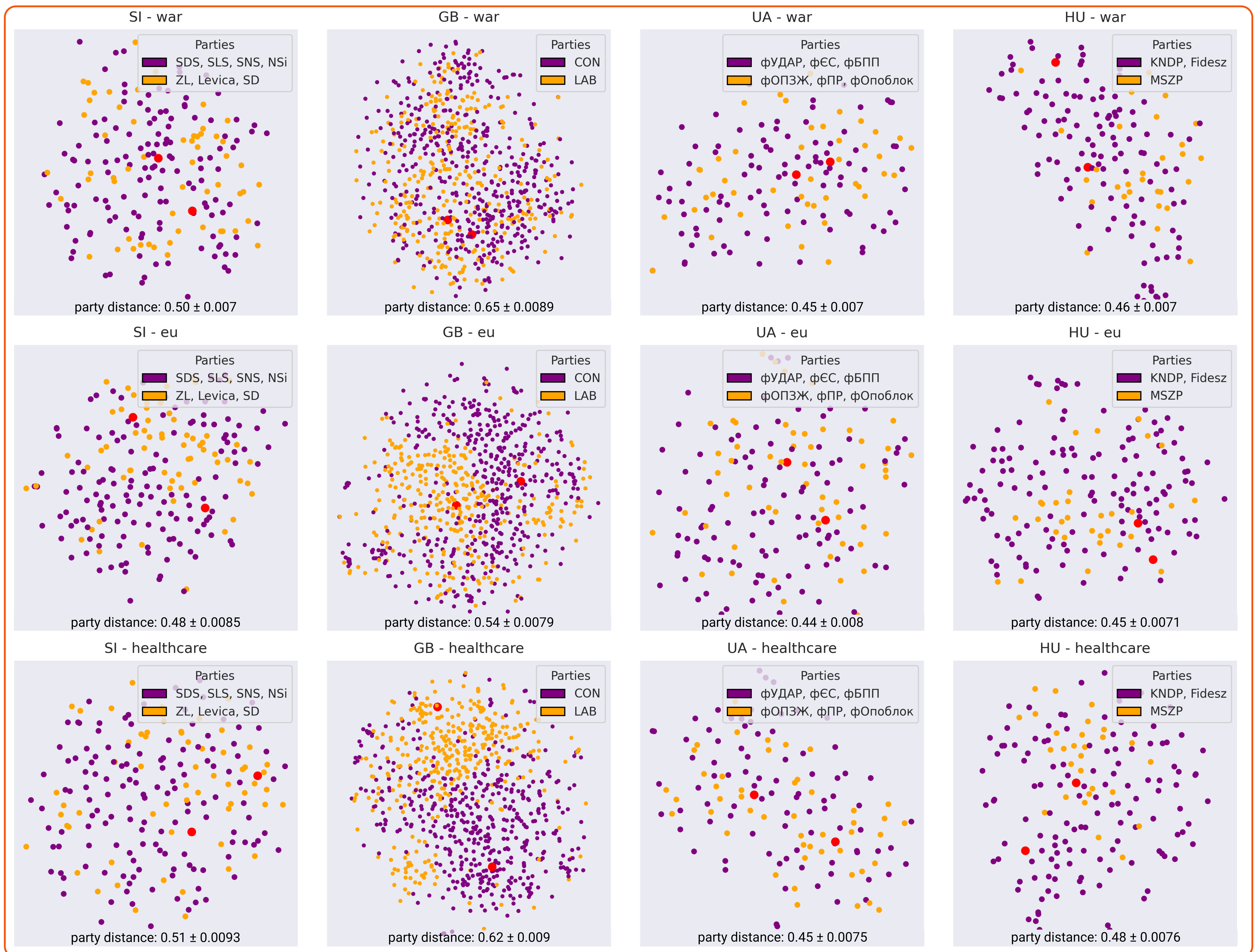
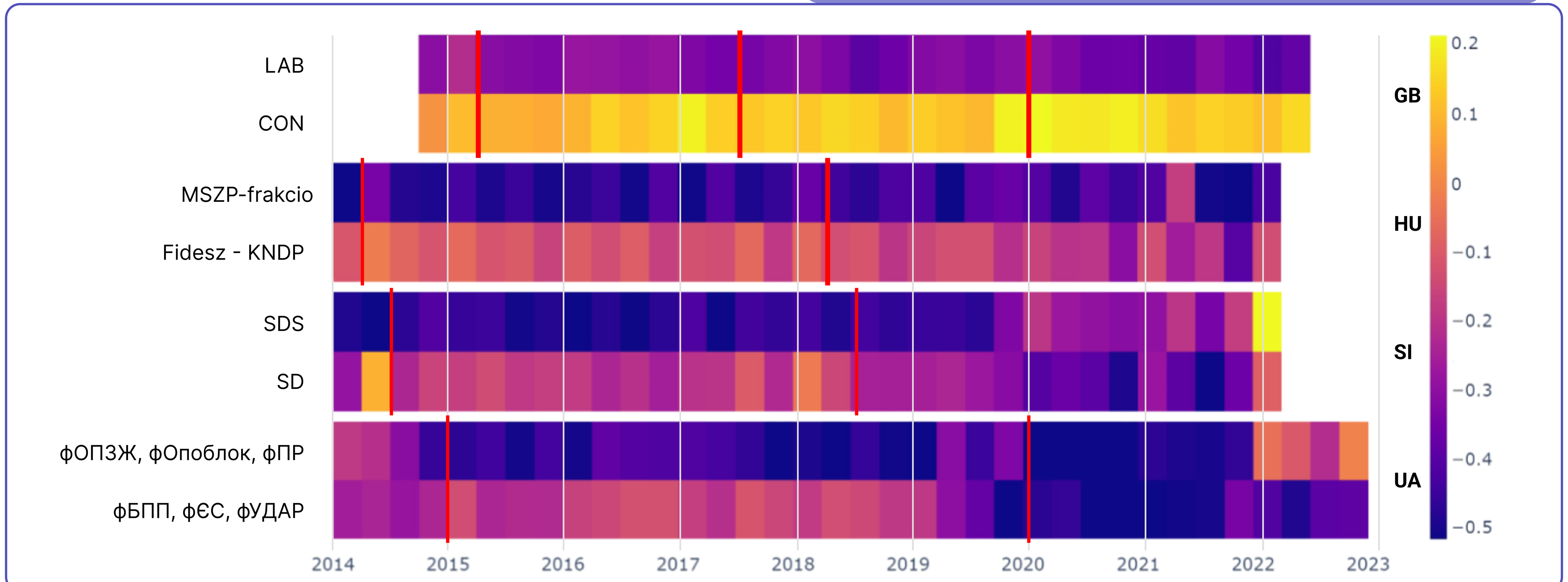


Figure 1:
Visualizing Embeddings
Scatterplots are showcasing parliamentarians from opposing groups of each parliament across the four thematic subcorpora.

Figure 2:
Temporal Sentiment Analysis
Heatmaps comparing opposing groups across the four parliaments. Vertical red lines represent elections.



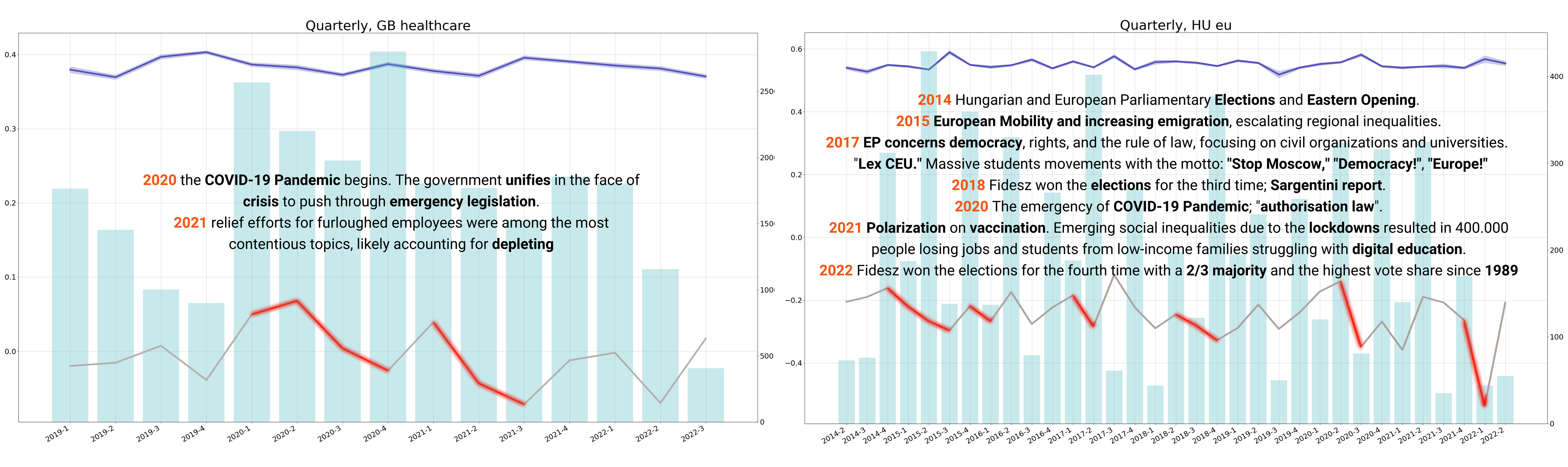


Figure 3:
Cosine Differences and Sentiments
Numerical difference of embedded opposition and coalition speeches (blue); average sentiment (red); and quarterly volume of speeches.

Challenges

- **Data** – the different volumes of the thematic subcorpora limit the interpretability of the results and likely affect the generalizability of the sentiment model and reproducibility of the topic modeling.
- **Methods** – The LLM-based approaches offer suggestive insights into polarization. However, variations in topic interpretability hindered the qualitative analysis of detected topics within themes.
- **Qualitative Analysis** – A principled revision of the qualitative post-analysis framework would increase confidence in the validity of our methods. Having only one language expert for Hungarian and Ukrainian slowed down this analysis.

Future research

- **Ontology** – crafting a more concrete definition of polarization to aid meaningful feature selection/identification and post-analysis.
- **Qualitative post-analysis for unsupervised methods** – the principled incorporation of corpus-based approaches in post-analysis could positively impact reliability.
- **Standardization of computational methodology.**
- **LLMs and Human readability** – Inclusion of explainable AI and white-box methods to aid interpretation and post-analysis.

DARIAH-FI · University of Helsinki · Aalto University · FIN-CLARIN · DARIAH-EU · CLARIN · HIIT HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY