

In partial fulfilment of the requirements for the degree of
Masters of Sciences in Business Analytics
Central European University, Department of Economics

Visualizing Political Narratives: A Machine Learning and Knowledge Graph Approach

Ghadena Hgaig

Under the Supervision of Eduardo De La Rubia and Aron Fellegi

June 2025

Vienna, Austria

Table of Content

Author's declaration.....	3
Copyright Notice.....	4
Public Summary.....	5

Author's declaration

I, the undersigned, **Ghadena Hgaig**, candidate for the MSc degree in Business Analytics declare herewith that the present thesis titled “Visualizing Political Narratives: A Machine Learning and Knowledge Graph Approach” is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography.

I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright.

I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

Vienna, 07 June 2025

Ghadena Hgaig

Copyright Notice

Copyright © Ghadena Hgaig, 2025. Visualizing Political Narratives: A Machine Learning and Knowledge Graph Approach.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

To view a copy of this license, visit: <https://creativecommons.org/licenses/by-nc-nd/4.0/>



[1] CC License icon by Creative Commons, <https://creativecommons.org/about/downloads/>

Public Summary

In today's fast-paced media environment, political campaign managers face two critical challenges: efficiently identifying politically relevant news articles from a large volume of media coverage, and visualizing the evolving network of political actors, institutions, and events. Traditional keyword-based filtering techniques often lead to high false-positive rates, while manual curation methods are accurate but not scalable. Large Language Models (LLMs) can improve text filtering but are computationally expensive, creating a need for scalable, cost-effective solutions.

This project addressed these challenges by developing a two-part solution:

- (1) an efficient machine learning classifier based on non-LLM features to pre-filter politically relevant articles, and
- (2) an interactive knowledge graph to map political discourse across the United States, France, and Germany.

The classifier was trained on a dataset of 6,303 articles covering the 2024 U.S. Presidential Election, the 2024 French General Election, and the 2025 German Federal Election. Using XGBoost and an engineered set of non-LLM features, the model achieved a recall rate of 81% in identifying politically relevant articles. Further improvements were observed when focusing on English full-text articles, where recall increased to 96%. Fragility tests confirmed that the model was robust even when top predictive features were removed.

Once filtered, the relevant articles were processed by an LLM to extract named entities and relationships. These entities and relationships formed the basis of a dynamic knowledge graph, offering a connected view of the political media landscape. Centrality analysis showed that U.S. political discourse was heavily personality-driven, with figures like Donald Trump dominating media narratives. In contrast, France and Germany exhibited more institution-centered discourse, emphasizing coalitions and parties rather than individuals. Community detection revealed nuanced structures, such as the presence of echo chambers and sentiment gaps. For instance, the Kamala Harris community in the U.S. graph exhibited high internal friendliness but significant external hostility, suggesting a strategic need for campaign managers to engage broader audiences.

While the project demonstrates the potential of combining machine learning and knowledge graphs for political media analysis, it also faces limitations. The dataset reflects media narratives rather than public opinion, and relationship extraction depends on explicit mentions in news articles, missing implicit or social media-based relationships. Despite these limitations, the system offers a scalable, cost-effective, and insightful tool for political campaign managers to filter thousands of articles efficiently and gain deeper insights into media discourse.