

Report for the Python project:

Analyzing political ideology in German Parliamentary speeches using LLMs

Project and Report are prepared by Ulyana Tsurkanu, Jan Dzemiashkevich, Fandi Chen

Course: [Module- Coding \[Python \]](#)

Professor: [Stanisław Giziński](#)

Research Goals

INTRO. Political scientists often study politicians and parties by looking at their positions on key issues. As Benoit and Laver (2006, p.13) note, it's hard to understand political debates without thinking about where people "stand" on issues. These positions let us compare how close or far apart different actors are ideologically. For references, we used Party Policy in Modern Democracies (Benoit & Laver, 2006) ([link](#)) and Codebook LLMs: Evaluating LLMs as Measurement Tools for Political Science Concepts (Haltermann & Keith, 2025)

This project aims to study patterns of ideology in speeches from the German Bundestag using computational text analysis. Specifically, it seeks to:

- Identify the main ideological orientations of different parties.
- Compare how similar or different parties are based on what they say.
- Test whether modern AI methods (like embeddings) can reveal meaningful ideological patterns, while treating LLM-based analysis as an exploratory measurement tool rather than a definitive coding scheme (Haltermann & Keith, 2025).

By thinking about ideology in terms of positions and relationships, this study follows a long tradition in political science that treats politics as competition and disagreement within an ideological space.

This project addresses the following research questions:

1. Do Bundestag speeches reflect recognizable party ideologies?
2. Are parties that are ideologically closer more similar in their speech content?
3. Can modern embedding-based methods capture ideological structure beyond simple left-right scales?

Data Sources

For this study, we used the Bundestag-v2 dataset (Threite, 2023) obtained from [Hugging Face](#). The dataset contains a large collection of speeches from the German Bundestag, with each row representing a single speech. The primary unit of analysis is the speech text, which is labeled with the speaker's political party.

Data Characteristics:

- Language: German.
- Format: Parquet Size: 143,000 speeches in the training split.
- Text: The full text of the speech, varying from short procedural statements to long, detailed policy discussions.
- Party: The political party of the speaker (categorical, 8 values, including CDU/CSU, SPD, FDP, Bündnis 90/Die Grünen, and PDS/Die Linke)
- The unit of analysis is a single speech.

Our study examines publicly observable political positions rather than private opinions, which aligns with Benoit and Laver (2006, p.18): *we can only observe what politicians say publicly*.

Methodology

Ideology as Multidimensional

- Political views are rarely one-dimensional. Most political disagreements require several dimensions to describe properly (Benoit & Laver, 2006, p.16). So our project looks at ideology across multiple categories rather than just a simple left-right scale.

Classifying Ideology

- Each speech was analyzed using a large language model, which assigned probabilities across six ideological categories: Conservative, Liberal, Social Democratic, Green, Left-wing Socialist, Nationalist/Right-wing Populist
- Probabilities sum to 1, so each speech can express elements of multiple ideologies rather than being put into a single box.

Aggregating by Party

-To understand party-level ideology, scores were averaged across all speeches from the same party. *As noted by Halterman and Keith (2025), aggregation can mask classification uncertainty at the individual text level, so party-level results should be interpreted as approximate patterns rather than precise estimates*

Embedding-Based Analysis and Visualization

- Each speech was also converted into a high-dimensional vector (embedding) using Sentence-BERT. These embeddings capture the meaning of the text. Principal Component Analysis (PCA) then reduced these vectors to two dimensions so we could visualize them.

Note: This two-dimensional view is a simplification, used for exploration rather than a definitive map of ideology.

Measuring Similarity Between Parties

- Cosine similarity was calculated between party-level ideological profiles. Parties that were very similar were connected in a network, showing how closely their speeches align ideologically. This emphasizes relationships between parties rather than exact positions.

Results

Party Ideologies. The results show clear ideological differences:

- CDU/CSU – mostly conservative
- FDP – mainly liberal
- SPD – social democratic
- Greens – green/eco-oriented
- PDS/Die Linke – left-wing socialist

These findings suggest that speeches reflect recognizable ideological patterns that match known party identities

Now we will look closer. Specifically discussing the graphs

The First graph. Average Ideological Profiles of Bundestag Parties

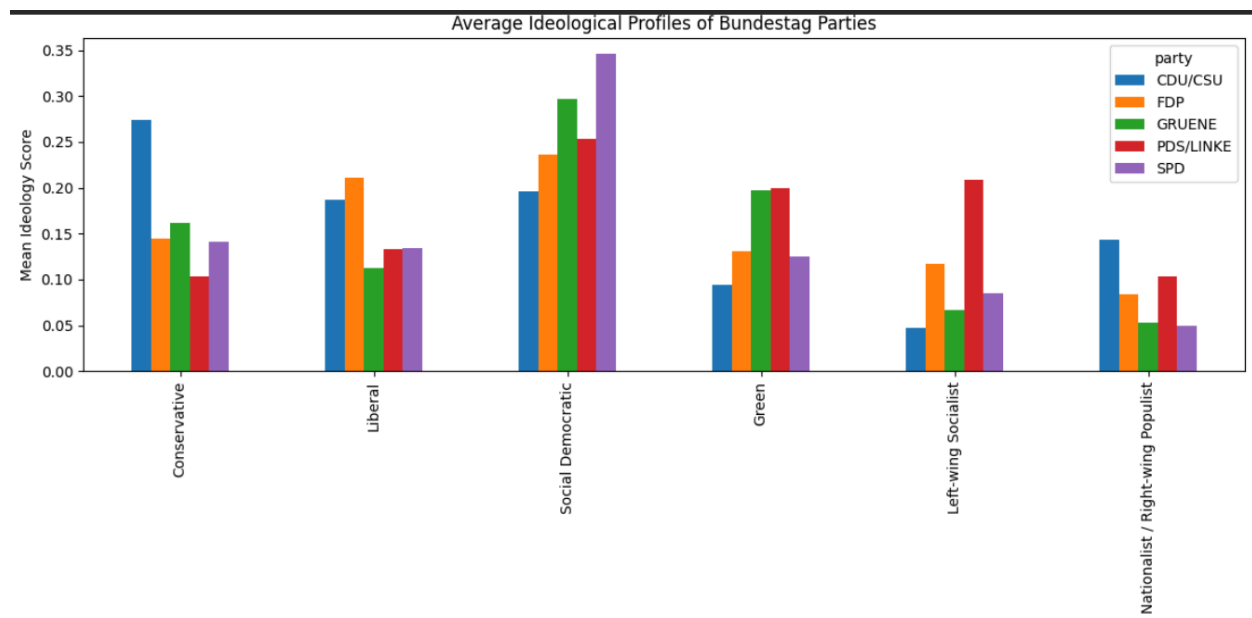
First Figure shows the average ideological positions of five major German political parties (CDU/CSU, FDP, Greens, PDS/Die Linke, and SPD) across six ideological categories.

The results clearly reflect the main political orientations of each party.

- CDU/CSU scores highest on the conservative dimension, confirming its position as a center-right party. It shows lower support for green and left-wing socialist ideas.

- FDP is strongest in the liberal category, which fits its focus on economic freedom and individual rights. Other ideological scores are moderate to low.
- The Greens (GRÜNE) stand out in the green dimension, while also showing strong social democratic and moderate left-wing socialist positions, reflecting their environmental and progressive agenda.
- PDS/Die Linke has the highest score in left-wing socialist ideology, as well as strong green and social democratic values, confirming its left-wing orientation.
- SPD scores highest in the social democratic category, which matches its traditional role as Germany's main social democratic party.

However, what we find interesting is that Parties are not mixed randomly. Speeches from the same party tend to appear close to each other. This means parties usually talk in a similar way and focus on similar ideas. Parties that are closer in ideology (for example SPD and the Greens) often overlap in the graph. This shows they sometimes talk about similar topics or use similar language. Parties like CDU/CSU and Die Linke are usually far from each other. This means their speeches are very different in content and ideology. Not all speeches fit perfectly. Some speeches from a party appear far from the main group of that party. This can happen because: the speech is about a special topic, the speaker uses a different tone, or the party is internally diverse.



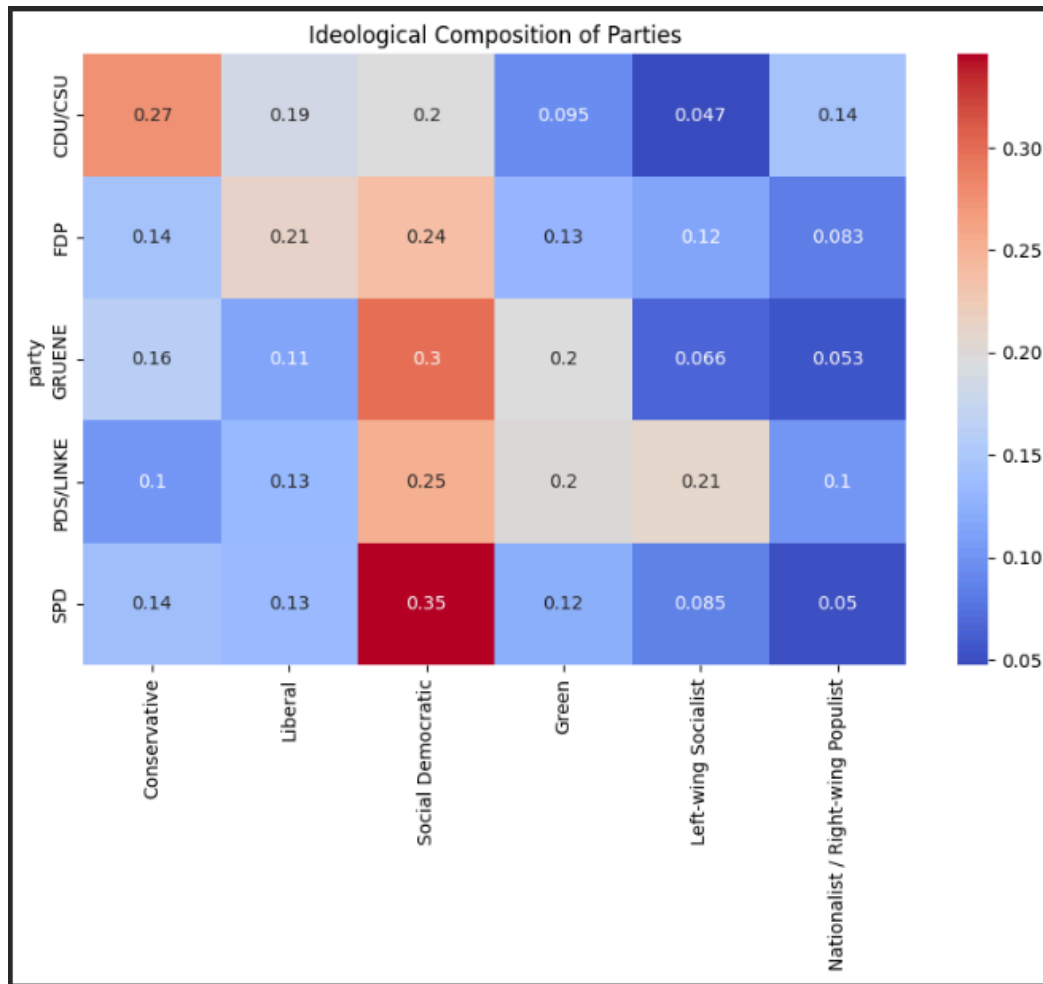
The Second graph. Ideological Composition of Parties

This heatmap illustrates the average ideological composition of parliamentary speeches for each party. Each row represents a party, while each column corresponds to an ideological category.

Warmer colors indicate a stronger presence of a given ideology in party discourse, whereas cooler colors indicate weaker representation.

- The CDU/CSU shows a clear dominance of conservative ideology, with additional but smaller liberal and social democratic elements. The presence of left-wing socialist and nationalist/right-wing populist content is relatively low. Overall, this pattern is consistent with the party's traditional center-right ideological position.
- The FDP displays its strongest associations with liberal and social democratic ideologies, while conservative and left-wing socialist elements are less pronounced. This reflects the FDP's emphasis on liberal economic and political principles combined with selective social policy themes.
- Bündnis 90/Die Grünen exhibits a strong concentration in social democratic and green ideological categories. Conservative and nationalist/right-wing populist elements are minimal. This distribution aligns closely with the party's environmental focus and progressive policy orientation.
- The PDS/Die Linke demonstrates high scores in social democratic, green, and especially left-wing socialist ideologies. Conservative and nationalist dimensions are weak. This profile clearly reflects the party's left-wing ideological stance and its emphasis on social justice and redistribution.
- The SPD shows the strongest social democratic orientation among all parties. It also contains moderate conservative and liberal elements, while nationalist/right-wing populist content remains very low. This suggests a broad center-left ideological profile that incorporates multiple ideological traditions.

Overall, the heatmap reveals that party ideologies are not expressed as single, pure categories. Instead, each party combines several ideological elements, with one or two dominant dimensions. This supports the idea that political ideology in parliamentary discourse is multidimensional rather than strictly divided along a single ideological axis.



The Third graph. Ideological Space of Bundestag Speeches

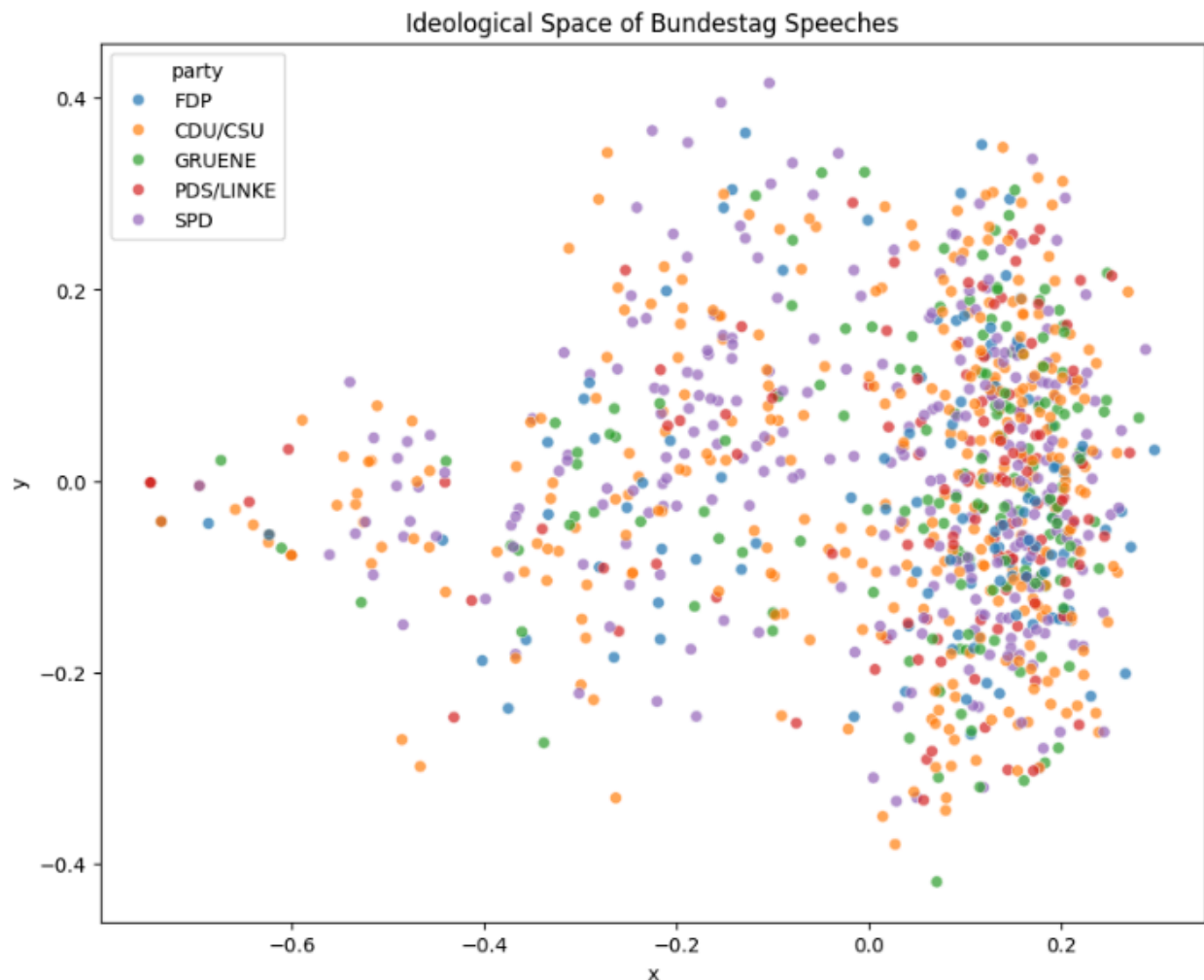
This figure visualizes the ideological space of Bundestag speeches based on sentence embeddings reduced to two dimensions using PCA. Each point represents a single speech, and colors indicate the party affiliation of the speaker. The distance between points reflects how similar or different speeches are in terms of language and content.

The graph shows that speeches are **not randomly distributed** but tend to form **loose clusters by party**. Speeches from the same party often appear close to each other, indicating consistent patterns in language use and thematic focus. At the same time, these clusters are not sharply separated, and there is noticeable overlap between parties, especially those that are ideologically closer.

This overlap suggests that parliamentary discourse shares common topics and vocabulary across parties, particularly on widely debated policy issues. However, parties such as CDU/CSU and

PDS/Die Linke tend to occupy more distant regions of the space, reflecting stronger ideological differences in how issues are framed.

Overall, the figure illustrates that ideological differences in parliamentary speeches exist but are **gradual rather than strictly divided**. This supports the view that political ideology is multidimensional and that party discourse often blends shared themes with party-specific perspectives.



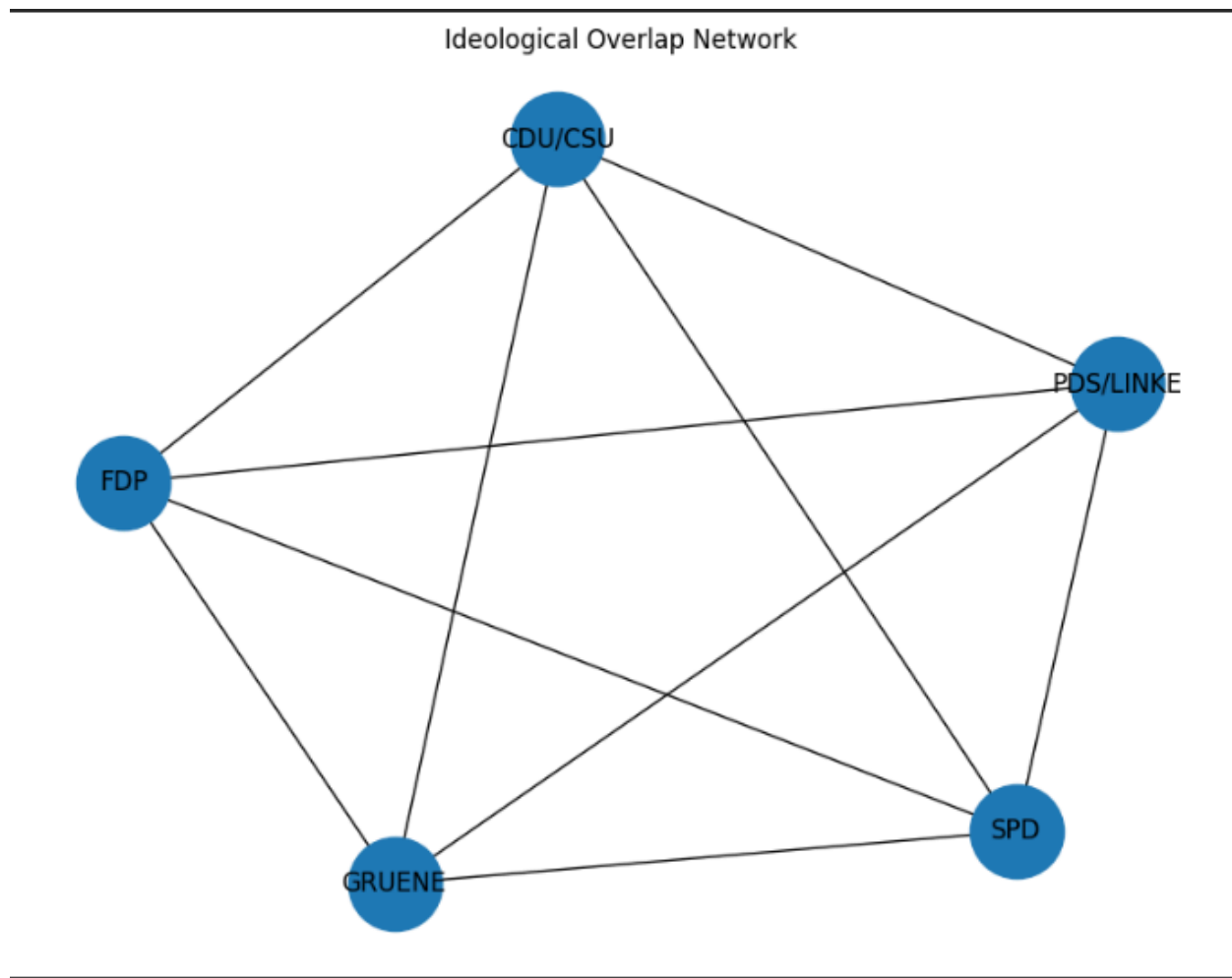
The Fourth graph. Ideological Overlap Network

This figure shows the ideological overlap network between Bundestag parties based on the similarity of their aggregated ideological profiles. Each node represents a political party, and connections between nodes indicate a strong ideological similarity above a chosen threshold. The thickness of the lines reflects the strength of similarity: thicker lines indicate greater ideological overlap, while thinner lines represent weaker overlap.

The network reveals that all parties are connected, suggesting that no party is completely isolated in ideological terms. This indicates that parliamentary discourse contains shared ideological elements across the political spectrum, even among parties with different core positions.

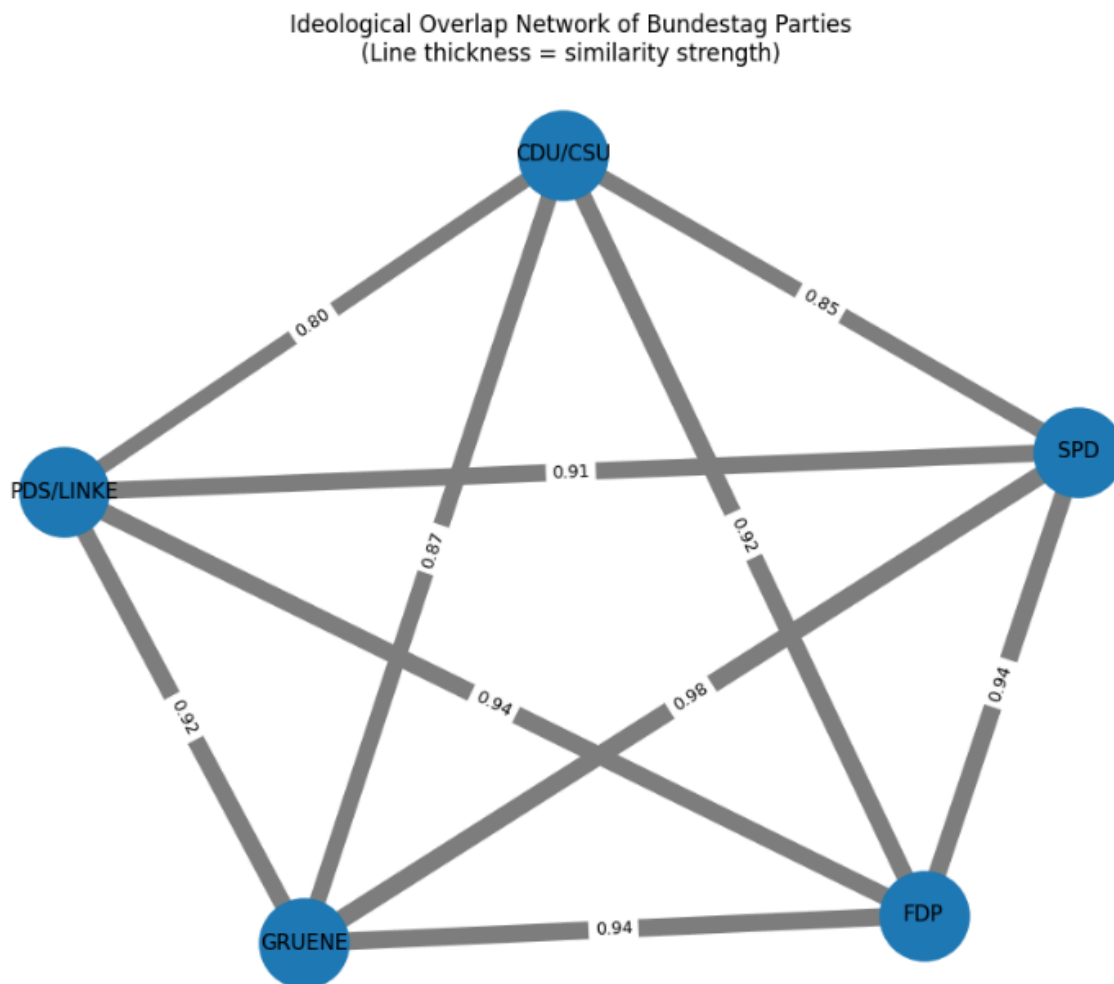
At the same time, the presence of multiple connections highlights varying degrees of ideological proximity. Parties that are closer in political orientation, such as SPD and Bündnis 90/Die Grünen, show strong overlap, while connections between more ideologically distant parties reflect partial but meaningful similarities in specific policy areas.

Overall, the network emphasizes that ideological relationships between parties are relational rather than absolute. Instead of existing in isolation, parties are embedded in a shared ideological space where overlap and distance coexist, reinforcing the multidimensional nature of political ideology in parliamentary discourse.



The Fifth graph. Ideological Overlap Network of Bundestag Parties

The figure illustrates the ideological similarity between major German Bundestag parties, where thicker lines represent stronger ideological overlap. Overall, the network shows a high level of similarity among all parties, with most connections being strong. The strongest overlaps are observed between FDP and the Greens, as well as between FDP and PDS/Die Linke, indicating close proximity in their average ideological profiles. SPD also demonstrates strong similarity with FDP and PDS/Die Linke, suggesting its central position within the ideological network. In contrast, CDU/CSU shows slightly weaker, though still substantial, connections with the other parties, reflecting its more distinct conservative orientation. In general, the graph highlights a dense and interconnected ideological landscape, where party positions partially overlap despite clear ideological differences.



Challenges and Limitations

Since the speeches are in German, model performance may depend on the quality of German-language representations in the underlying models.

- Simplifying complex political views into a few categories loses some nuance.
- AI-based classification can have biases depending on the model and prompts.
- Aggregating at the party level may hide internal differences within parties.
- Any multidimensional analysis involves choices; there's no single "true" way to map ideology (Benoit & Laver, 2006, p.40).
- LLM-based text classification provides approximate measurements, and classification uncertainty at the speech level may affect aggregated results (Halterman & Keith, 2025).

References

- Benoit, K., & Laver, M. (2006). *Party policy in modern democracies*.
- Threite, J. (2023). *Bundestag-v2 dataset*.
- Halterman, L., & Keith, D. (2025). *Codebook LLMs: Evaluating LLMs as Measurement Tools for Political Science Concepts*.
- OpenAI ChatGPT (2023). Used as part of methodology for ideological classification and text analysis.