Graph-Augmented LLMs for Swiss MP Ideology Prediction

Yifei Yuan[1], Luis Salamanca[1], Sophia Schlosser[2], Laurence Brandenberger[2]*
[1]Swiss Data Science Center, ETH Zürich, Zürich, Switzerland
[2]Department of Political Science, University of Zürich, Zürich, Switzerland
yifei.yuan@sdsc.ethz.ch, laurence.brandenberger@ipz.uzh.ch

Abstract

Approximating the ideological position of Members of Parliament (MPs) is a fundamental task in political science, helping researchers understand legislative behavior, party alignment, and policy preferences. While Large Language Models (LLMs) have shown promising results in estimating MPs ideological stances, there are more actors and elements in the parliamentary system, and relations between them, that could provide a wider and more informative picture. However, due to the complexity of integrating them in the prediction task, these additional elements are generally ignored. In this work, we propose an LLM framework, *PG-RAG*, that implements a retrieval-augmented generation pipeline: it first queries a political knowledge graph (KG) and then integrates the resulting graph-structured information into the context. This allows for capturing both textual semantics and inter-MP relationships, another relevant information source in any parliamentary system. We evaluate the approach on the task of ideology prediction, using data from a Swiss parliamentary dataset. When comparing graph-augmented models against several state-of-the-art baselines, the results demonstrate that incorporating this enriched information, which encodes information about different entities and relations, improves prediction performance. These results help to highlight the value of domain-specific relational information in modeling political behavior.

1 Introduction

A central question in political science is how to infer the ideological positions of Members of Parliament (MPs) from their observable political behavior (Poole and Rosenthal, 1985; Clinton et al., 2004). Scholars have increasingly used text-based methods to estimate ideology scores from speeches, parliamentary debates, and manifestos, enabling more fine-grained and scalable assessments of MPs' political positions (Laver et al., 2003; Slapin and Proksch, 2008; Proksch and Slapin, 2010; Lauderdale and Herzog, 2016).

As Large Language Models (LLMs) have demonstrated strong performance across a range of NLP tasks, recent studies have explored their potential for predicting MPs ideological stances from textual sources, leveraging their ability to capture semantic nuances and latent political signals embedded in parliamentary speeches (Liu et al., 2022; Bernardelle et al., 2024). Although these approaches have shown reasonable performance, several limitations persist. First of all, they typically treat MPs as independent text generators and ignore the relational structure in parliamentary systems, such as co-sponsorship networks, committee memberships, or party blocs, despite recent work suggesting that such relational information can substantively enrich MPs' representations (Russo et al., 2023). Moreover, LLM-based methods often struggle in low-data settings or when long-term dependencies across multiple parliamentary sessions must be considered, highlighting the need for methods that can jointly leverage textual content and structured relational knowledge (Huang et al., 2024).

To address these limitations, we propose the integration of graph-structured information to capture ideological alignment and relational influence. By using graph-augmented LLMs, which combine the language understanding capabilities of LLMs with structured relational knowledge captured in graphs, we investigate whether this additional information can improve prediction accuracy. Specifically, we propose a **Political Graph Retrieval-Augmented Generation** (*PG-RAG*) framework, where a graph encoding parliamentary relationships – such as co-sponsorship links, committee memberships, party affiliations,

*Corresponding author.

and ideological clusters – is queried to retrieve relevant subgraphs for each MP. To encode the complexity of the retrieved graphs, comprised by nodes and relations connected to certain MPs and related to specific parliamentary aspects, we explored two approaches. First, we leverage the great summarization capabilities of LLMs and prompt the model to first summarize the elements retrieved. Second, we explore the ability of the LLM to understand the retrieved graph by providing it as a raw set of nodes and relations. For each case, the additional context, either the obtained summary or the raw graph, is provided as additional information to a pre-defined prompt, also including some general metadata of the MP.

The goal of our work is to investigate whether graph-structured information indeed improves ideology prediction and which types of relational signals are most informative. To assess this, we compare the described approaches against strong baselines, covering several state-of-the-art LLMs. To provide a more systematic evaluation, we benchmark zero-shot and few-shot LLM setups, as well as models of different size. Our experimental results show that incorporating graph information improves prediction performance, with the effect being particularly pronounced for smaller-scale LLMs. We also observe that LLM-based models struggle when positioning Social Democrats, highlighting directions for future analysis.

2 Related Work

2.1 Ideology Prediction

Ideology prediction aims to infer the political or ideological orientation of individuals, groups, or textual content and has been widely studied in the Political Science and NLP domains. Early work focused on predicting the ideological leanings of political actors, such as legislators or parties, using legislative speeches, manifestos, or voting records (e.g., Cox and Poole, 2002; Bakker et al., 2015; Kraft et al., 2016; Vafa et al., 2020; Patil et al., 2019). Approaches from the Political Science domain relied traditionally on scaling procedures (Poole and Rosenthal, 1985; Slapin and Proksch, 2008; Burnham, 2024). In contrast, some NLP-based approaches relied on traditional machine learning models and linguistic features, including SVMs (Sapiro-Gheiler, 2019) and RNNs (Sinno et al., 2022), to distinguish ideological positions. More recent studies have leveraged pre-trained language models to capture richer contextual representations for ideology prediction across different domains and languages; for instance, Liu et al. (2021) pre-train a Transformer-based language generator to minimize ideological bias in generated text. With the emergence of LLMs, researchers have begun examining whether ideological orientations can be inferred directly from generated or summarized content, as well as how biases present in training data may affect model predictions (Liu et al., 2022; Bernardelle et al., 2024; Kim et al., 2025).

2.2 LLMs for Political Tasks

Recent studies have explored the capabilities of LLMs in political analysis. Prior work shows that LLMs can perform tasks such as political stance detection (Li et al., 2021; Wagner et al., 2024; Pangtey et al., 2025), ideology classification (Haroon et al., 2025), policy analysis (Chen et al., 2025), often achieving performance comparable to or surpassing traditional NLP models. Researchers have also examined the extent to which LLMs encode political biases or ideological patterns in their training data, investigating whether model outputs reflect systematic political preferences or framing effects (Zhang, 2025; Kim et al., 2025; Rettenberger et al., 2024). In addition, several studies evaluate LLMs in political reasoning and multimodal settings, including tasks such as policy debate generation (Dzeparoska et al., 2023; Chuang et al., 2025), argument analysis (Li et al., 2025), and political question answering (Santurkar et al., 2023).

3 Our Framework

3.1 Preliminary

Given a dataset D consisting of k MP records, $D = \{(i, p_i, g_i, l_i)\}_{i=1}^{k}$, where p_i denotes the party of the i-th MP, g_i denotes the corresponding party group, and l_i represents the ideology score, the ideology prediction task aims to learn a function F that maps an specific MP i, and its party and group information (i, p_i, g_i) to the corresponding ideology score l_i.

$$l_i = F(i, p_i, g_i) \tag{1}$$

The information of the MP encoded in i can simply contain personal and demographic data, such as age and education, which can already support the task of ideology detection. However, more

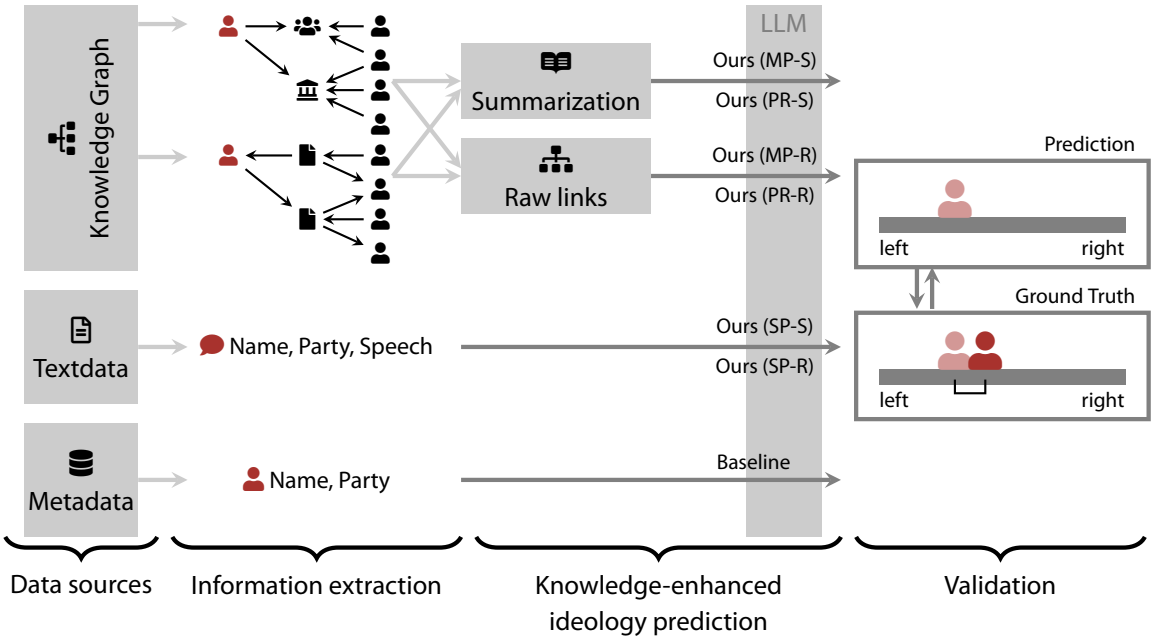

Figure 1: The overall framework of our proposed method *PG-RAG*.

complex information can be additionally provided. In the following sections, we discuss the proposed methodology *PG-RAG* to further enrich i with information queried from a Political KG.

3.2 PG-RAG Method

As shown in Figure 1, we propose a RAG-inspired methodology that leverages information extracted from a political knowledge graph (KG) to address the task of ideology prediction. Specifically, the approach uses the information contained in a (1) **political KG**, on which it performs (2) **subgraph extraction**, to finally carry out (3) **knowledge-enhanced ideology prediction**. The following sections detail these steps.

3.2.1 Political KG

The Political KG utilized is built using the information extracted from the Bulletin of the Swiss Parliament, as detailed in (Salamanca et al., 2024). The schema implemented by this KG aims at encoding the policy-making process, from the moment a pursuit text is proposed by a committee, to all the discussions occurring in the parliament chambers related to it. This is captured through entities such as `Pursuit` and `Speech`, with relations encoding temporal dependencies. Furthermore, rich metadata, related to the MPs, Parties, etc., is additionally integrated into the graph, providing further context. A subset of this Political KG, which corresponds to the legislative periods 48th to 51st, ranging from 2007 to 2023, is available at (Brandenberger et al., 2024), with further details on the KG structure and its usage.

3.2.2 Subgraph extraction

Due to the large size and complexity of the KG, we decided to define meaningful subgraphs that can be queried independently when generating additional context for the prediction tasks. Each of these subgraphs comprised a subset of entities and the relations connecting them, linked to a specific parliamentary process. Now, given an MP record, we first match it to the corresponding `Person` node in the KG. Starting from this MP node, the three subgraphs explored are defined according to the following paradigms:

- **Speech-centric (SP)**: We assume that an MP's speeches and legislative activities provide important signals of their ideological position. Therefore, we collect all speeches linked to the MP `Person` node, through the relation `gives`. This serves as an important **baseline** for the other two scenarios, as the retrieved subgraph contains only a single relation and is purely textual.

- **MP-centric (MP)**: In this setting, we extract a subgraph that captures the structural and institutional relationships surrounding the MP (see Figure 6). The extracted subgraph includes entities representing the MPs political

affiliations and institutional roles, such as the `Party`, `Parliamentary Group`, `Committee`, and `Chamber` to which the MP belongs or is elected. In addition, the subgraph incorporates contextual entities describing administrative and geographic connections, such as the represented `Canton`.

- **Pursuit-centric (PR)**: The pursuit-centric paradigm focuses on the legislative activities initiated or supported by the MP. Specifically, starting from the target MP node, we retrieve all `Pursuit` entities that are sponsored or co-sponsored by the MP (see Figure 7). These pursuits represent legislative proposals or initiatives that reflect the MP's policy interests and political priorities.

Above, the subgraphs are ordered by their complexity. First, in the SP subgraph, the only relevant information is the textual data contained in the `Speech` node. Hence, this approach is similar to recent methods relying on the semantics of textual data. On the contrary, the MP subgraph captures a true graph structure by querying different entity types within the 1-hop vicinity of the `Person` node. Finally, the PR subgraph increases the complexity by enabling 2-hop extraction, as well as entities connected by different relation types. The specific queries used to parse the graph are provided in the Appendix D.

3.2.3 Knowledge-enhanced ideology prediction

Given the extracted subgraph, we need to generate a suitable representation that can be provided as additional context to the LLMs, aiming at improving the ideology prediction task. We propose the following two approaches to encode the subgraph:

- **Summarization (S)**: Motivated by (Zhao et al., 2023), we first serialize the subgraph into natural language sentences. Specifically, each triplet in the knowledge graph is converted into a textual statement that describes the relationship between entities. These serialized statements collectively form a structured textual representation of the MP's political context, including institutional affiliations, legislative activities, and other relevant relations captured in the subgraph. We use GPT-5 to summarize the subgraph, using the prompts presented in Appendix A.3.

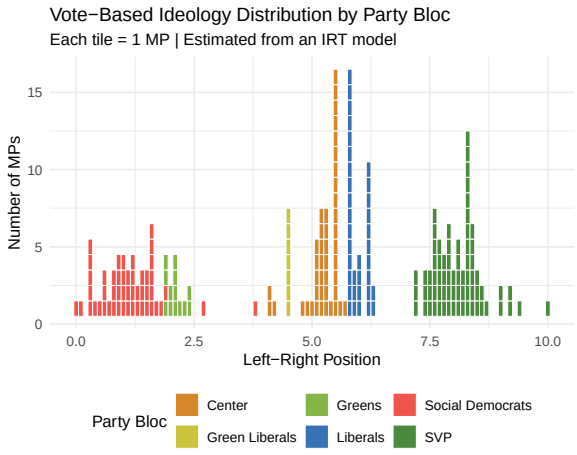

Figure 2: Vote-based ideology scores of Swiss members of the National Council

- **Raw-Graph (R)**: The graph is provided as retrieved from Neo4J (Neo4j Core Team, 2024). The JSON formatting encodes the connected nodes as sub-elements, indicating explicitly the relationship type and the different attributes' values. An example is provided in Appendix C.

Specifically, SP-S provides a summarization of all the queried speeches. In contrast, SP-R presents a subset of them independently. Similarly, (MP/PR)-(S/R) encode truly relational information between different entity types, either serialized into text, or provided as a raw-graph. We report the performance of these 6 different methods accordingly in the following section.

4 Experiments

4.1 Experimental Setup

Dataset. We collect a dataset from the Swiss National Council, the lower chamber of the Swiss Federal Assembly, comprising 225 unique members of parliament (MPs) during the 50th legislative period (2015-2019). The number exceeds the 200 seats ($N = 200$) because some MPs left and were replaced over the four-year period. We compare ideology predictions for these MPs to vote-based scaled estimates. These vote-based estimates stem from a random sample of 1000 votes recorded in the Swiss National Council. All votes in the National Council are recorded electronically and include points of order votes, standard votes on proposals as well as final votes. The

Method	MAE ↓	MSE ↓	RMSE ↓	RC ↑
GM	2.29	7.44	2.73	-
PM	0.30	0.20	0.44	0.97
PBM	0.33	0.23	0.48	0.97
zero-shot				
GPT-5	0.75	1.06	1.03	0.94
Qwen3-8B	1.20	2.32	1.52	0.86
Qwen3-32B	1.11	2.03	1.43	0.86
Apertus-8B	2.01	4.58	2.14	0.77
PG-RAG (MP-S)	**0.72**	0.77	0.88	0.94
PG-RAG (MP-R)	0.73	**0.74**	**0.86**	**0.94**
few-shot				
GPT-5	0.61	0.61	0.78	0.94
Qwen3-8B	0.88	1.57	1.25	0.90
Qwen3-32B	0.87	1.45	1.21	0.90
Apertus-8B	2.47	7.91	2.81	0.43
PG-RAG (MP-S)	**0.58**	**0.58**	**0.76**	**0.94**
PG-RAG (MP-R)	0.61	0.60	0.78	0.94

Table 1: Main results comparing all baseline methods to our best performing models, those using the MP-centric subgraph. RC represents Ranking Correlation.

Subgraph	Enc.	MAE	MSE	RMSE	RC
zero-shot					
MP	S	**0.72**	0.77	0.88	0.94
SP	S	0.73	0.78	0.88	0.93
PR	S	0.83	1.03	1.02	0.93
MP	R	0.73	**0.74**	**0.86**	**0.94**
SP	R	0.76	0.78	0.89	0.94
PR	R	0.79	0.89	0.94	0.93
few-shot					
MP	S	**0.58**	**0.58**	**0.76**	**0.94**
SP	S	0.60	0.70	0.84	0.93
PR	S	0.68	0.83	0.91	0.94
MP	R	0.61	0.60	0.78	0.94
SP	R	0.62	0.59	0.76	0.94
PR	R	0.62	0.60	0.77	0.94

Table 2: Exhaustive results for all variants of PG-RAG, using different subgraphs and context encoding methods. It is important to highlight that the SP method does not provide a graph per-se, but rather a summarization of all speeches (SP-S), or a subset of complete speeches (SP-R).

voting data is provided by the Swiss Parliament[1] and incorporated into the DemocraSci KG (Brandenberger et al., 2026). We use a widely-used dimensional-reduction technique, a two-parameter Item Response Theory model, to estimate a one-dimensional model, as per the standard approach applied in Political Science (for a methodological discussion of scaling procedures, see Cox and Poole 2002; Cai et al. 2016; Bailey and Voeten 2018, based on early scaling techniques developed by Poole and Rosenthal 1985). We use the `mirt`-package (Chalmers, 2012) in the Statistical Environment `R` to estimate vote-based ideology scores. Figure 2 shows the stacked distribution of ideology scores (one-dimensional, commonly interpreted as left-right ideological positions).

Evaluation Metrics. We evaluate the ideology prediction performance using the following evaluation metrics: (1) **Regression Metrics**: Since the task involves predicting continuous ideology scores, we measure prediction accuracy using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Regular Mean Squared Error (RMSE). (2) **Ranking Metrics**: We also assess whether the model preserves the relative ordering of ideological positions. For this purpose, we employ Spearman's rank correlation (ρ). These sets of metrics provide complementary perspectives, helping to reach more insightful and interpretable results.

Compared Methods. We compare our method against several baselines, including: (1) **Naive baselines**: Global Mean (GM), which assigns each MP the overall mean ideology score; Party Mean (PM), which assigns each MP the average ideology score of their party; and Party Bloc Mean (PBM), which assigns each MP the mean ideology score of their party bloc and can be considered an **upper bound** for this category of methods. (2) LLM-based methods: We also compare our method with several state-of-the-art LLMs, including GPT-5 (Singh et al., 2025), the Swiss LLM Apertus (Apertus et al., 2025) (8B version), Qwen3-8B and Qwen3-32B-AWQ (Yang et al., 2025), under both zero-shot and few-shot settings. For the few-shot setup, we randomly select three examples from the vote-based dataset to serve as in-context demonstrations (prompts see Appendix A.1 and A.2). For all of them, we use the default parameter settings.

4.2 Experimental Results

Table 1 reports the performance of different methods on the ideology prediction task. Several observations can be made: among LLM-based baselines, GPT-5 achieves the best zero-shot performance (MAE = 0.75, Rank Corr = 0.94), while Qwen3-32B performs moderately well. In contrast, Apertus-8B shows substantially weaker performance, suggesting that general-purpose LLMs struggle to infer ideology reliably without structured signals. When incorporating graph-derived summary knowledge, our approach improves pre-

[1] https://www.parlament.ch/de/ratsbetrieb/abstimmungen/abstimmungs-datenbank-nr

diction accuracy. In the zero-shot setting, the MP-centric subgraph reduces MAE to 0.72, while maintaining the same ranking correlation. Furthermore, the MSE is significantly reduced from 1.06 to 0.74 for the PG-RAG (MP-R) case, which demonstrates how our approach is capable of reducing the prediction error even in cases where GPT-5 deviates substantially from the ground truth value. For the few-shot scenario, i.e., when we provide in-context MP examples in the prompt, with their associated metadata and ideology score, the results are more on par. Still, the PG-RAG (MP-S) approach provides some slight improvement. Nevertheless, it is important to highlight that, during our experiments, we noticed a really brittle behavior of the few-shot approach, and adding more examples did not always lead to better results.

Overall, these results demonstrate that injecting structured knowledge distilled from political graphs into LLM prompts substantially enhances ideology prediction, enabling LLMs to better capture ideological ordering among MPs. In particular, the improvements in MSE for the zero-shot scenario allow to demonstrate how the proposed methods can help recover from predictions that deviate substantially from the ground truth value.

In Table 2, we provide results for all the subgraphs queried and used as context. As discussed before, both SP cases resemble previous approaches in which only some textual input is provided as context. In all scenarios, the MP-centric subgraph provides the best results. We believe this is because the MP-centric subgraph provides rich complementary information, such as committees and chambers, while maintaining a moderate context size. On the contrary, the results for the Pursuit-centric subgraph present a degraded performance, likely related to its larger size and complexity, which the LLM still falls short in correctly leveraging. We added more detailed study in Appendix E.

5 Extensive Analysis

5.1 Party-wise Analysis

Figure 3 plots the best LLM-based ideology prediction (MP-S) against the vote-based ideology scores. MPs are colored by party blocs (with representatives of the Christian and Conservative Democratic Parties merged into the Center bloc). Overall, the two ideology scores correlate at a score of $r = 0.963$ (Pearson's, $t = 53.904$, 223 degrees of freedom, $p - value < 0.0001$). The three smaller centric party blocs all show strong coherence in their ideology predictions. The two pole party blocs (dominated by SVP and SP members) show more dispersion around the diagonal, indicating that the LLM sometimes predicts the MPs to be more left or right-leaning than would be expected from their voting behavior. MPs from the Green party bloc (represented by members of the Green party and affiliate communist parties) are generally predicted to be more left by the LLM. This more left-leaning prediction stems from the fact that the Green party is often ideologically positioned to the left of the Social Democrats and is renowned for collaborating with the Swiss communist parties (who inhabit the left-extreme position) (Ladner, 2019, 2012; Hug and Schulz, 2007; Jolly et al., 2022). However, in the 50th legislative period, the Greens have often voted along left ideological positions and deviated towards the center in order to strengthen their alliances to left-leaning centric members. This has brought the MPs from the green party away from the extreme-left position in terms of their voting behavior. The strongest deviation in the LLM prediction stems from the Social Democrats. Here, two factors are at work. First, the LLM judges Social Democrats to be more right-leaning in their ideology than they present in their voting behavior. Second, the LLM

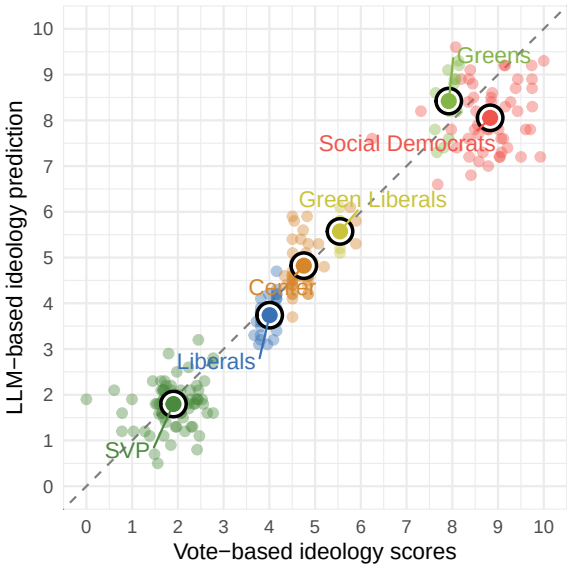

Figure 3: Scatterplot of the best LLM prediction scores (MP-S) vs. the vote-based ideology scores. MPs are colored by party blocs and within-bloc averages are highlighted (black-lined circles).

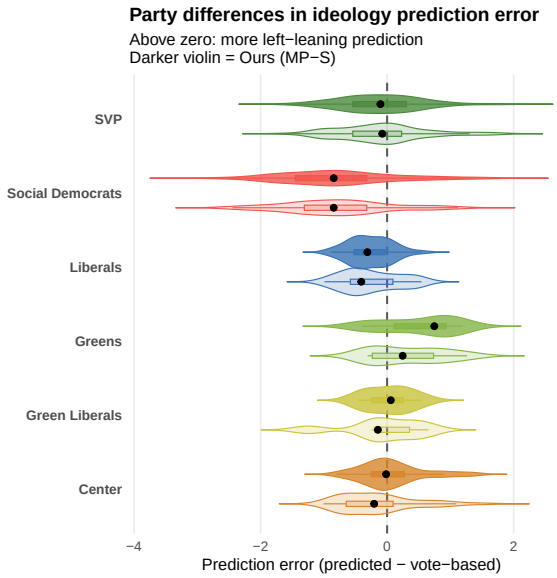

Figure 4: Prediction error by party blocs. The dark violins represent PG-RAG (MP-S) model, the lighter violins show the GPT-5 model.

does not deal with the within-party diversity as well as it does with other party blocs. It is worth investigating whether this dispersion stems from the fact that the LLM is unsure where to place these MPs, or whether the signals from their affiliations are different to their signals from their voting behavior. The latter could possibly stem from increased party discipline in voting, however, then the ideological placements of Social Democrats from these votes would be more unified.

Next, we compare the model predictions from GPT-5 (few-shot) and PG-RAG (MP-S) by party bloc in order to see where the additional graph information has helped improve model predictions. Figure 4 presents prediction errors. For some party blocs, the additional graph information helps contract the predictions around the true (vote-based) values. This is the case for the Liberals, the Green Liberals as well as for the Center party bloc. However, for others, the raw GPT-5 predictions are closer to the vote-based ideologies (Greens, SVP). This should not necessarily be interpreted as a failure of our model, but rather that the inclusion of additional MP information has shifted the ideology predictions. It is well-known that voting behavior is not the only ideological indicator in legislative studies (Snyder Jr and Groseclose, 2000; Rheault and Cochrane, 2020; Barber, 2022) and, as we have indicated, it is based on legislative behavior that is biased in and of itself (e.g., Carrubba et al., 2006, 2008; Hug, 2010). As such, it would be interesting to study in larger and more varied datasets whether the LLM-based and graph-enhanced ideology predictions reflect a more nuanced ideology placement that encompasses MPs' legislative behavior outside voting.

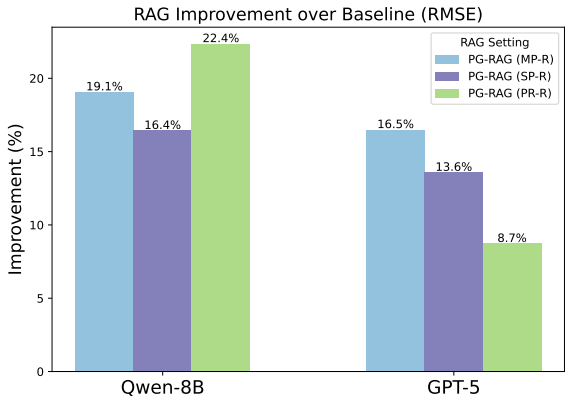

Figure 5: Performance improvements from RAG over the non-RAG baseline for Qwen-8B and GPT-5.

Figure 4 also shows the difficulties the LLM-based models face when placing Social Democrats. Here, the prediction errors span the largest range in both models, indicating that both the raw GPT-5 as well as the graph-enhanced predictions are generally more right-leaning for Social Democrats than their voting behavior would suggest. The reason for this distortion needs to be explored further: is it based on heterogeneous actions by these MPs that make them difficult to pinpoint, or is it an inherent bias in LLMs to bias Social Democrats more towards the right?

5.2 Backbone Analysis

Figure 5 shows the RMSE improvement from RAG over the non-RAG baseline for Qwen-8B and GPT-5 under the zero-shot scenario, with the context encoding method as raw subgraph (R). From the figure, we observe that the PG-RAG (MP-R) setting yields the most significant improvement for GPT-5, reducing the error rate by 16.5%. This suggests that for large-scale models, providing high-density MP-centric graph context is highly effective. Additionally, Qwen-8B consistently shows higher percentage improvements across all RAG settings (ranging from 16.4% to 22.4%) compared to GPT-5 (8.7% to 16.5%). This indicates that graph-based context provides a more substantial "knowledge boost" to the model with a smaller scale than to the more advanced GPT-

Person	Party (Party Bloc)	Summary of the MP-centric subgraph	GT	GPT-5	PG-RAG
MP-A	LDP (Liberals)	MP-A (b. [YEAR], [CITY]) is a National Council (NR) member for [CANTON]. He belongs to the Liberal Democratic Party (LDP) and sits with the FDP-Liberale parliamentary group (RL), the liberal bloc in Swiss politics. He serves on the [COMMITTEE SEAT]. The LDP/FDP family in Switzerland generally emphasizes a market-oriented economy, individual liberties, and strong support for education, research, and cultural policy.	4.2	4.6	4.2
MP-B	FDP-Liberale (Liberals)	MP-B (b. [YEAR], [CITY]) is a National Council (NR) member representing the canton of [CANTON]. He is a member of FDP. Die Liberalen and sits in the FDP-Liberal parliamentary group (formerly Freisinnig-demokratische Fraktion). The FDP in Switzerland is associated with market-oriented economic policy, competitiveness and individual freedoms. He serves on the [COMMITTEE SEAT A] and the [COMMITTEE SEAT B] committees, focusing on these policy areas.	3.8	2.9	3.6
MP-C	SVP (SVP)	MP-C (born [YEAR] in [CITY]; citizen of [CITIZENSHIP]) represents [CANTON]] in the Council of States (SR). He is a member of the Swiss People's Party (SVP) and its parliamentary group (V). His committee work includes the [COMMITTEE SEAT] and an ad hoc committee [COMMITTEE NUMBER]. The SVP is a right-wing, conservative party emphasizing national sovereignty, restrictive immigration policy, lower taxes, and skepticism toward EU integration.	1.7	2.1	1.8
MP-D	SP (Social Democrats)	MP-C (born [YEAR] in [CITY]) is a National Council member for the canton of [CANTON] from the Sozialdemokratische Partei der Schweiz (SP) and sits in the Sozialdemokratische Fraktion. The SP is a center-left social-democratic party advocating social justice, strong public services, labor rights, and progressive social policy. Naef serves on the [COMMITTEE SEAT A] and the [COMMITTEE SEAT B], indicating a focus on international and legal matters. He is a citizen of the city of [CITIZENSHIP].	9.0	6.5	7.1

Table 3: Examples (de-identified) of the MP ideology prediction from GPT-5 and PG-RAG (MP-S). The summaries are generated through GPT-5 using the information queried from the KG, for the MP-centric subgraph. The prompt used to carry out ideology prediction contained the fully identified information.

5. Interestingly, while the PG-RAG (PR) (Pursuit-based) setting was the strongest for Qwen-8B, it shows the weakest improvements for GPT-5, providing an 8.7% improvement. This suggests that smaller models depend more on external structured guidance, whereas larger models can internally absorb and reason over the same information. Overall, integrating graph knowledge provides a moderate, stable improvement for both models, acting as a reliable middle-ground strategy.

5.3 Case Study

We further demonstrate several examples to understand our model, MP-S case, compared to GPT-5 raw model in Table 3. We find that across the cases, our model's predictions consistently align more closely with the ground truth than GPT-5. For MP-A and MP-B, both members of liberal parties, our prediction is almost identical to the GT. Similarly, for MP-C, our prediction is nearer to the GT than GPT-5's. In these cases, our method is better able to leverage the graph and textual context to make accurate predictions. In addition, we observe that for politicians from left-leaning parties, such as MP-D, our model predicts a score of 7.1, which is closer to the ground-truth value of 9.0 compared to GPT5's prediction of 6.5. This example highlights the challenge of accurately predicting left-leaning MPs, consistent with our findings in Section 5.1. This difficulty arises because left-leaning MPs often have high ideology scores, while LLMs tend to generate more moderate predictions. Our model partially mitigates this difficulty, producing predictions closer to the ground truth for some of these cases. By leveraging textual summaries from the knowledge graph, which capture rich contextual information, such as committee memberships and policy focus, our model enables better predictions.

6 Conclusion

We introduce *PG-RAG*, a RAG-inspired graph-augmented LLM framework for political ideology prediction. Leveraging the data from a Swiss parliamentary knowledge graph, we explore three subgraph scenarios, speech-centric, MP-centric, and pursuit-centric, with two context encoding methods for each, correspondingly. We then compare our method with several LLMs. Our experiments show that with graph-structured relational data, our approach captures the complex web of inter-MP relationships and parliamentary elements that define legislative behavior. In addition, with graph knowledge, the model shows improved understanding of political tendencies across different parties. These results lay the groundwork for future work on extending to additional parties and enhancing performance on left-leaning parties.

Limitations

While our model demonstrates strong overall predictive performance, it still exhibits reduced accuracy when predicting positions of left-leaning parties. This suggests that the model may not fully capture the nuances in the rhetoric or policy preferences characteristic of these groups. Additionally, our current dataset and analysis are limited to a subset of MPs, and extending the model to a broader set of representatives could improve generalizability. However, this expansion is not feasible for Switzerland due to data constraints and the limited availability of annotated parliamentary records.

Acknowledgments

This work is supported by the Swiss National Science Foundation (grant 10.003.190, measuring political success).

References

Project Apertus, Alejandro Hernández-Cano, Alexander Hägele, Allen Hao Huang, Angelika Romanou, Antoni-Joan Solergibert, Barna Pasztor, Bettina Messmer, Dhia Garbaya, Eduard Frank urech, Ido Hakimi, Juan García Giraldo, Mete Ismayilzada, Negar Foroutan, Skander Moalla, Tiancheng Chen, Vinko Sabolec, Yixuan Xu, Michael Aerni, and 84 others. 2025. Apertus: Democratizing open and compliant llms for global language environments. *Preprint*, arXiv:2509.14233.

Michael A Bailey and Erik Voeten. 2018. A two-dimensional analysis of seventy years of united nations voting. *Public Choice*, 176(1):33–55.

Ryan Bakker, Catherine De Vries, Erica Edwards, Liesbet Hooghe, Seth Jolly, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Anna Vachudova. 2015. Measuring party positions in europe: The chapel hill expert survey trend file, 1999–2010. *Party Politics*, 21(1):143–152.

Michael Barber. 2022. Comparing campaign finance and vote-based measures of ideology. *The Journal of Politics*, 84(1):613–619.

Pietro Bernardelle, Leon Fröhling, Stefano Civelli, Riccardo Lunardi, Kevin Roitero, and Gianluca Demartini. 2024. Mapping and influencing the political ideology of large language models using synthetic personas. *Companion Proceedings of the ACM on Web Conference 2025*.

Laurence Brandenberger, Julian Minder, Luis Salamanca, Sophia Schlosser, Lilian Gasser, Vincent Jung, Kourosh Shariat, Marta Balode, Anna Schmidt-Rohr, Leon Babi, Fernando Perez-Cruz, and Frank Schweitzer. 2024. DemocraSci - a parliamentary Knowledge Graph (4 legislative periods).

Laurence Brandenberger, Sophia Schlosser, Luis Salamanca, Lilian Gasser, Marta Balode, Julian Minder, Vincent Jung, Yaren Durgun, Leon Babic, Fernando Perez-Cruz, and Frank Schweitzer. 2026. DemocraSci: A Knowledge Graph on the Swiss parliament. *Submitted Manuscript*.

Michael Burnham. 2024. Semantic scaling: Bayesian ideal point estimates with large language models. *arXiv preprint arXiv:2405.02472*.

Li Cai, Kilchan Choi, Mark Hansen, and Lauren Harrell. 2016. Item response theory. *Annual Review of Statistics and Its Application*, 3(1):297–321.

Clifford Carrubba, Matthew Gabel, and Simon Hug. 2008. Legislative voting behavior, seen and unseen: A theory of roll-call vote selection. *Legislative Studies Quarterly*, 33(4):543–572.

Clifford J Carrubba, Matthew Gabel, Lacey Murrah, Ryan Clough, Elizabeth Montgomery, and Rebecca Schambach. 2006. Off the record: Unrecorded legislative votes, selection bias and roll-call vote analysis. *British Journal of Political Science*, 36(4):691–704.

R Philip Chalmers. 2012. mirt: A multidimensional item response theory package for the r environment. *Journal of statistical Software*, 48:1–29.

Yuxin Chen, Peng Tang, Weidong Qiu, and Shujun Li. 2025. Using llms for automated privacy policy analysis: Prompt engineering, fine-tuning and explainability. *ArXiv*, abs/2503.16516.

Yun-Shiuan Chuang, Ruixuan Tu, Chengtao Dai, Smit Vasani, Binwei Yao, Michael Henry Tessler, Sijia Yang, Dhavan Shah, Robert Hawkins, Junjie Hu, and 1 others. 2025. Debate: A large-scale benchmark for role-playing llm agents in multi-agent, long-form debates. *arXiv preprint arXiv:2510.25110*.

Joshua Clinton, Simon Jackman, and Douglas Rivers. 2004. The statistical analysis of roll call data. *American Political Science Review*, pages 355–370.

Gary W Cox and Keith T Poole. 2002. On measuring partisanship in roll-call voting: The us house of representatives, 1877-1999. *American Journal of Political Science*, pages 477–489.

Kristina Dzeparoska, Jieyu Lin, Ali Tizghadam, and Alberto Leon-Garcia. 2023. Llm-based policy generation for intent-based management of applications. In *2023 19th International Conference on Network and Service Management (CNSM)*, pages 1–7. IEEE.

Muhammad Haroon, Magdalena Wojcieszak, and Anshuman Chhabra. 2025. "whose side are you on?" estimating ideology of political and news content using large language models and few-shot demonstration selection. In *IJCNLP-AACL*.

Wenyu Huang, Guanchen Zhou, Mirella Lapata, Pavlos Vougiouklis, Sébastien Montella, and Jeff Z. Pan. 2024. Prompting large language models with knowledge graphs for question answering involving long-tail facts. *Knowl. Based Syst.*, 324:113648.

Simon Hug. 2010. Selection effects in roll call votes. *British Journal of Political Science*, 40(1):225–235.

Simon Hug and Tobias Schulz. 2007. LeftRight positions of political parties in Switzerland. *Party Politics*, 13(3):305–330.

Seth Jolly, Ryan Bakker, Liesbet Hooghe, Gary Marks, Jonathan Polk, Jan Rovny, Marco Steenbergen, and Milada Anna Vachudova. 2022. Chapel hill expert survey trend file, 1999–2019. *Electoral studies*, 75:102420.

Junsol Kim, James Evans, and Aaron Schein. 2025. Linear representations of political perspective emerge in large language models. *ArXiv*, abs/2503.02080.

Peter E. Kraft, Hirsh Jain, and Alexander M. Rush. 2016. An embedding model for predicting roll-call votes. In *EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings*.

Andreas Ladner. 2012. Switzerland's green liberal party: a new party model for the environment? *Environmental politics*, 21(3):510–515.

Andreas Ladner. 2019. Switzerland: The" green" and" alternative parties". In *New Politics In Western Europe*, pages 155–165. Routledge.

Benjamin E Lauderdale and Alexander Herzog. 2016. Measuring political positions from legislative speech. *Political Analysis*, 24(3):374–394.

Michael Laver, Kenneth Benoit, and John Garry. 2003. Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(02):311–331.

Hao Li, Viktor Schlegel, Yizheng Sun, Riza Theresa Batista-Navarro, and Goran Nenadic. 2025. Large language models in argument mining: A survey. *ArXiv*, abs/2506.16383.

Yingjie Li, Tiberiu Sosea, Aditya Sawant, Ajith Jayaraman Nair, Diana Inkpen, and Cornelia Caragea. 2021. P-stance: A large dataset for stance detection in political domain. In *Findings*.

Ruibo Liu, Chenyan Jia, Jason Wei, Guangxuan Xu, Lili Wang, and Soroush Vosoughi. 2021. Mitigating political bias in language models through reinforced calibration. *ArXiv*, abs/2104.14795.

Yujian Liu, Xinliang Frederick Zhang, David Wegsman, Nick Beauchamp, and Lu Wang. 2022. Politics: Pretraining with same-story article comparison for ideology prediction and stance detection. *ArXiv*, abs/2205.00619.

Neo4j Core Team. 2024. Neo4j - the worlds leading graph database.

Lata Pangtey, Anukriti Bhatnagar, Shubhi Bansal, Shahid Shafi Dar, and Nagendra Kumar. 2025. Large language models meet stance detection: A survey of tasks, methods, applications, challenges and future directions. *ArXiv*, abs/2505.08464.

Pallavi Patil, Kriti Myer, Ronak Zala, Arpit Singh, Sheshera Mysore, Andrew McCallum, Adrian Benton, and Amanda Stent. 2019. Roll call vote prediction with knowledge augmented models. In *CoNLL 2019 - 23rd Conference on Computational Natural Language Learning, Proceedings of the Conference*.

Keith T Poole and Howard Rosenthal. 1985. A spatial model for legislative roll call analysis. *American Journal of Political Science*, 29(2):357–384.

Sven-Oliver Proksch and Jonathan B Slapin. 2010. Position taking in European Parliament speeches. *British Journal of Political Science*, 40(03):587–611.

Luca Rettenberger, Markus Reischl, and Mark Schutera. 2024. Assessing political bias in large language models. *Journal of Computational Social Science*, 8.

Ludovic Rheault and Christopher Cochrane. 2020. Word embeddings for the analysis of ideological placement in parliamentary corpora. *Political Analysis*, 28(1):112–133.

Giuseppe Russo, Christoph Gote, Laurence Brandenberger, Sophia Schlosser, and Frank Schweitzer. 2023. Helping a friend or supporting a cause? disentangling active and passive cosponsorship decisions in the u.s. congress. *ACL2023, The 61st Annual Meeting of the Association for Computational Linguistics. Toronto, Canada, July 2023.*

Luis Salamanca, Laurence Brandenberger, Lilian Gasser, Sophia Schlosser, Marta Balode, Vincent Jung, Fernando Perez-Cruz, and Frank Schweitzer. 2024. Processing large-scale archival records: The case of the Swiss parliamentary records. *Swiss Political Science Review*, 30(2):140–153.

Shibani Santurkar, Esin Durmus, Faisal Ladhak, Cinoo Lee, Percy Liang, and Tatsunori Hashimoto. 2023. Whose opinions do language models reflect? *ArXiv*, abs/2303.17548.

Eitan Sapiro-Gheiler. 2019. Examining political trustworthiness through text-based measures of ideology. In *AAAI Conference on Artificial Intelligence*.

Aaditya K. Singh, Adam Fry, Adam Perelman, Adam Tart, Adithya Ganesh, Ahmed El-Kishky, Aidan McLaughlin, Aiden Low, AJ Ostrow, Akhila Ananthram, Akshay Nathan, Alan Luo, Alec Helyar, Aleksander Madry, Aleksandr A Efremov, Aleksandra Spyra, Alex Baker-Whitcomb, Alex Beutel, Alex Karpenko, and 464 others. 2025. Openai gpt-5 system card.

Barea M. Sinno, Bernardo Oviedo, Katherine Atwell, Malihe Alikhani, and Junyi Jessy Li. 2022. Political ideology and polarization: A multi-dimensional approach. In *North American Chapter of the Association for Computational Linguistics*.

Jonathan B Slapin and Sven-Oliver Proksch. 2008. A scaling model for estimating time-series party positions from texts. *American Journal of Political Science*, 52(3):705–722.

James M Snyder Jr and Tim Groseclose. 2000. Estimating party influence in congressional roll-call voting. *American Journal of Political Science*, 44(2):193–211.

Keyon Vafa, Suresh Naidu, and David M Blei. 2020. Text-based ideal points. *arXiv preprint arXiv:2005.04232*.

Stefan Sylvius Wagner, Maike Behrendt, Marc Ziegele, and Stefan Harmeling. 2024. The power of llm-generated synthetic data for stance detection in online political discussions. *ArXiv*, abs/2406.12480.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, Chujie Zheng, Dayiheng Liu, Fan Zhou, Fei Huang, Feng Hu, Hao Ge, Haoran Wei, Huan Lin, Jialong Tang, and 41 others. 2025. Qwen3 technical report.

Tianyi Zhang. 2025. Probing political ideology in large language models: How latent political representations generalize across tasks. In *Conference on Empirical Methods in Natural Language Processing*.

Jianan Zhao, Le Zhuo, Yikang Shen, Meng Qu, Kai Liu, Michael M. Bronstein, Zhaocheng Zhu, and Jian Tang. 2023. Graphtext: Graph reasoning in text space. *ArXiv*, abs/2310.01089.

A LLM prompts

A.1 Zero-shot setting

```
Predict the ideology score of this Swiss MP
(0=Right, 10=Left). Important guidelines:
- Base your judgment primarily on the MP's
background information.
- Do NOT rely only on the party or bloc label.
- MPs within the same party or bloc can have
different ideological positions.
- Choose a precise value (e.g., 3.7 or 6.2).

Name: {MP Name}, Party: {MP Party}, Bloc: {MP
Party Block}
Return ONLY the number.
Score:
```

A.2 Few-shot setting

```
You are a political scientist. Below are
examples of MPs and their ideology scores (0 =
Far-Right, 10 = Far-Left):
Name: {example MP name} | Party: {example MP
party} | Party Bloc: {example MP party bloc} |
Score: {example MP ideology score}
...

Predict the ideology score of this Swiss MP
(0=Right, 10=Left). Important guidelines:
- Base your judgment primarily on the MP's
background information.
- Do NOT rely only on the party or bloc label.
- MPs within the same party or bloc can have
different ideological positions.
- Choose a precise value (e.g., 3.7 or 6.2).

Name: {MP Name}, Party: {MP Party}, Bloc: {MP
Party Bloc}
Return ONLY the number.
Score:
```

A.3 Prompt for subgraph summarization

```
You are a political science expert.

Given structured information about a Member of
Parliament (MP), write a concise neutral
description of this politician that could help
infer their political ideology.

Avoid speculation and keep the description
factual.

MP information:
{mp_context}

Write a short summary (around 500 characters)
describing the MP's political background and
potential ideological positioning.
```

B Subgraph Demonstration

B.1 MP-centric

Figure 6 shows the MP-centric subgraph. For the focal MP, we extract the chamber they are elected to, the committees they sit on, the parties and parliamentary groups they belong to, the canton they represent and the city they live in. Whenever relations are time-stamped we extract only those relations that are within the 50th legislative period.

B.2 Pursuit-centric

Figure 7 shows the pursuit-centric subgraph. This two-hop subgraph starts with the focal MP and their links to pursuits. These relations either represent sponsorship (i.e., submitting agent) or cosponsorship (i.e., supporting agent). The second hop represents links from the pursuit to other cosponsors as well as to committees or parliamentary groups who can act as sponsors themselves.

Model	MP-centric				Speech-centric				Pursuit-centric			
	MAE	MSE	RMSE	RC	MAE	MSE	RMSE	RC	MAE	MSE	RMSE	RC
zero-shot												
PG-RAG (r-10)	0.75	0.81	0.90	0.94	0.80	0.88	0.94	0.94	0.85	0.99	1.00	0.92
PG-RAG (r-50)	0.77	0.83	0.91	0.94	0.77	0.81	0.90	0.94	0.80	0.92	0.96	0.93
PG-RAG (r-100)	0.79	0.85	0.92	0.93	0.78	0.84	0.92	0.94	0.80	0.91	0.95	0.93
few-shot												
PG-RAG (r-10)	0.55	0.55	0.74	0.94	0.61	0.60	0.78	0.94	0.66	0.72	0.85	0.93
PG-RAG (r-50)	0.60	0.61	0.78	0.94	0.60	0.59	0.77	0.94	0.68	0.72	0.85	0.93
PG-RAG (r-100)	0.59	0.59	0.77	0.94	0.61	0.61	0.78	0.93	0.68	0.74	0.86	0.94

Table 4: Performance comparison of ideology prediction under different settings.

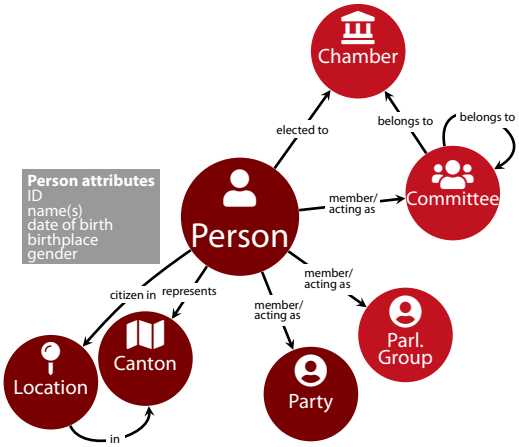

Figure 6: Graph demonstration of the MP-centric subgraph.

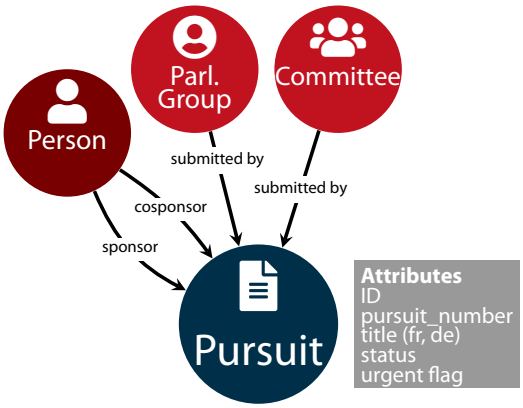

Figure 7: Graph demonstration of the pursuit-centric subgraph.

C Example Subgraph Demonstration

Here's an example of how our extracted Neo4J subgraph looks in the MP-centric scenario:

```
[
  {
    "csv_uid": "74",
    "original_label": null,
    "person_info": {
      "date_birth": "xxx",
      "uid": "xx",
      "gender": "x",
      "last_name": "xxx",
      "first_name": "xxx"
    },
    "graph_context": [
      {
        "type": "City",
        "rel": "BORN_IN",
        "properties": {
          "post_code": "4000",
          "name": "Basel"
        }
      },
      {
        "type": "City",
        "rel": "CITIZEN_IN",
        "properties": {
          "post_code": "4000",
          "name": "Basel"
        }
      },
      ...
    ]
  }
]
```

D Graph-parsing Queries

We provide the queries we used to parse the subgraph under the three scenarios.

D.1 Speech-centric

```
MATCH (p:Person {uid: $uid})
OPTIONAL MATCH (p)-[r]-(neighbor:Speech)
WHERE datetime(neighbor.time_end) >=
datetime("2015-11-30T00:00:00")
  AND datetime(neighbor.time_start) <=
  datetime("2019-12-01T23:59:59")
RETURN
    properties(p) as p_props,
    labels(neighbor)[0] as n_label,
    properties(neighbor) as n_props,
    type(r) as rel_type
```

D.2 MP-centric

```
 MATCH (p:Person {uid: $uid})-[r]-(neighbor)
WHERE any(label IN labels(neighbor) WHERE label
IN [
    'Chamber', 'Committee', 'Party', 'Canton',
```

```
    'Location', 'Parliamentary Group'
])
WITH p, neighbor, r,
     labels(neighbor)[0] AS n_label,
     properties(neighbor) AS n_props,
     type(r) AS rel_type
// Apply constraints for Chamber and Committee
WHERE (n_label = 'Chamber' AND rel_type =
'ELECTED_TO'
       AND r.date_election >= date("2015-11-30")
       AND r.date_election <= date("2019-12-01"))
   OR (n_label = 'Committee'
       AND r.date_joining >= date("2015-11-30")
       AND r.date_leaving <= date("2019-12-01"))
   OR (n_label <> 'Chamber' AND n_label <>
   'Committee')
RETURN DISTINCT properties(p) AS p_props,
n_label, n_props, rel_type
```

D.3 Pursuit-centric

```
MATCH (p:Person {uid: $uid})

OPTIONAL MATCH (p)-[r1]-(n1:Pursuit)

OPTIONAL MATCH (n1)-[rs:SUBMITTED_TO]->()
WHERE rs.date >= date("2015-11-30")
  AND rs.date <= date("2019-12-01")

OPTIONAL MATCH
(n1)-[r2:SPONSORS|COSPONSORS]-(n2)
WHERE n2 IS NULL OR n2 <> p

RETURN DISTINCT
    properties(p) AS p_props,
    labels(n1)[0] AS n1_label,
    properties(n1) AS n1_props,
    type(r1) AS r1_type,
    labels(n2)[0] AS n2_label,
    properties(n2) AS n2_props,
    type(r2) AS r2_type
```

E Model Variants Comparison

To avoid generating a massive context that could instead confound the model, we restrict the raw graph to N nodes. To ensure a correct coverage of the full subgraph, these are randomly selected among all the sub-elements. Table 4 shows the performance comparison under different setups when we randomly include N nodes with the prompt.