Hierarchical Classification of Propaganda Techniques in Slavic Texts in Hyperbolic Space

Christopher Brückner and Pavel Pecina
Charles University, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
{bruckner,pecina}@ufal.mff.cuni.cz

Abstract

Classification problems can often be tackled by modeling label hierarchies with broader categories in a graph and solving the task via node classification. While recent advances have shown that hyperbolic space is more suitable than Euclidean space for learning graph representations, this concept has yet to be applied to text classification, where node features first need to be extracted from text embeddings. A prototype of such an architecture is this contribution to the Slavic NLP 2025 shared task on the multi-label classification of persuasion techniques in parliamentary debates and social media posts. We do not achieve state-of-the-art performance, but outline the benefits of this hierarchical node classification approach and the advantages of hyperbolic graph embeddings.

1 Introduction

In times when large parts of the world forget history and fall victim to populist propaganda, now also spread through the internet, it becomes more important than ever to identify persuasion techniques in texts that aim to manipulate readers, rather than providing arguments of substance. However, their detection is a non-trivial task for language models and humans alike (Da San Martino et al., 2020). Propaganda appears in many forms and places:

The problem of classifying persuasion techniques in internet memes has previously been addressed as text-only and multimodal shared tasks at SemEval 2021 (Dimitrov et al., 2021) and SemEval 2024 (Dimitrov et al., 2024), the latter extending it to a multilingual hierarchical classification problem. Furthermore, SemEval 2023 (Piskorski et al., 2023) is concerned with multilingual news articles. This work is a contribution to the Slavic NLP 2025 Shared Task on the Detection and Classification of Persuasion Techniques in Slavic Languages (Piskorski et al., 2025), which extends the label taxonomy with additional persuasion techniques and focuses on five Slavic languages in two types of text: parliamentary debates in Bulgarian, Polish, Croatian, and Slovene, and social media posts in Russian.

In the following, we describe our approach to extending and embedding the label hierarchy defined by SemEval 2024. The problem is then tackled as a node classification problem, making use of hyperbolic geometry to improve the graph embeddings. Despite focusing on the classification of propaganda techniques, this methodology generalizes to related downstream tasks such as subject classification, given that the labels underlie a hierarchy or a hierarchy can be constructed. Rich subject hierarchies can be found, for example, in biomedical documents or testimonies of Holocaust survivors. Our implementation is available on GitHub[1].

2 Related Work

The architecture of the hierarchical text classification model closest to our system is HiAGM (Zhou et al., 2020). HiAGM projects text embeddings to node features and uses these as the input to a Hierarchy-GCN. This Hierarchy-GCN, inspired by Graph Convolutional Networks (Kipf and Welling, 2017), consists of three gated linear operations masked with three different adjacency matrices: parent-to-child direction, child-to-parent direction, and self-edges. In the parent-to-child direction, the edges are weighted based on transition priors estimated from the training data. Experiments in neural node classification and link prediction have shown that hyperbolic space is more suitable than Euclidean space for graph data modeling, especially when the graphs are more tree-like. There exist multiple equivalent models of hyperbolic space, and while the simpler and more interpretable Poincaré model is most commonly used, the hyperboloid model has been shown to be more numerically stable (Peng et al., 2022).

[1] https://github.com/chbridges/SlavHiTC

Proceedings of the 10th Workshop on Slavic Natural Language Processing (Slavic NLP 2025), pages 183–189
July 31, 2025

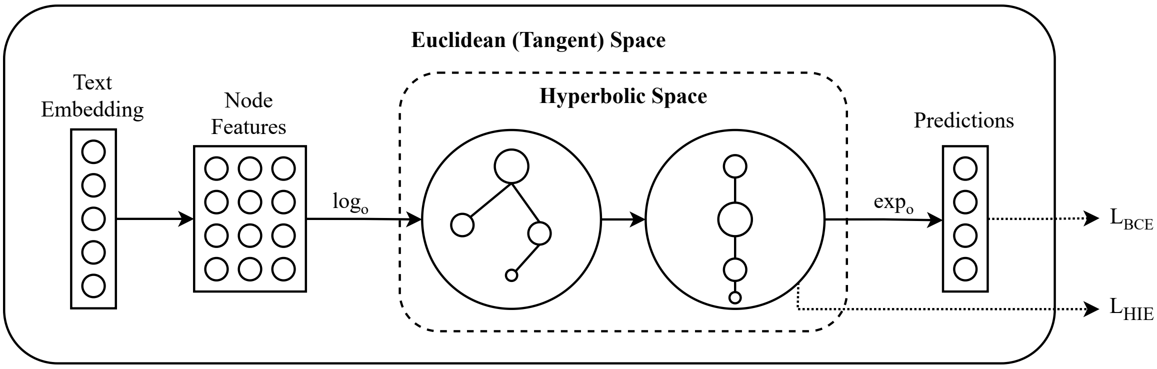

Figure 1: The complete model architecture. Node features are extracted from text embeddings, projected to a graph in hyperbolic space, and passed through graph convolutional layers. While the graph representations are optimized to use the bounded space more efficiently, the node features are projected to logits for binary predictions in Euclidean space. Larger nodes represent more general parent labels.

Hyperbolic Informed Embeddings (Yang et al., 2023) improve the embedding quality by equipping models such as Hyperbolic Graph Convolutional Networks (Chami et al., 2019) with an additional loss function to align the root node with the origin and stretch the remaining nodes across the hyperbolic space.

3 Methodology

In this section, we briefly introduce hyperbolic geometry before describing how we integrate it into the architecture of a text classification model. Utilizing label hierarchies in the classification head allows the model to leverage more contextual information and fall back to more general labels if it fails to predict the correct fine-grained label.

3.1 Poincaré model of hyperbolic space

The following abridged definition is based on Balazevic et al. (2019) and Peng et al. (2022).

The hyperbolic space is a Riemannian manifold $(\mathcal{M}, d)$ with constant negative curvature $-\kappa, \kappa > 0$. $\mathcal{M}$ can be locally approximated around $\mathbf{x} \in \mathcal{M}$ in the Euclidean *tangent space* $\mathcal{T}_\mathbf{x}\mathcal{M}$ via exponential and logarithmic maps $\exp_\mathbf{x}^\kappa \colon \mathcal{T}_\mathbf{x}\mathcal{M} \to \mathcal{M}$ and $\log_\mathbf{x}^\kappa \colon \mathcal{M} \to \mathcal{T}_\mathbf{x}\mathcal{M}$.

The n-dimensional Poincaré ball $(\mathbb{B}_\kappa^n, d^\mathbb{B})$ is defined by the open set and Riemannian metric

$$\mathbb{B}_\kappa^n = \left\{\mathbf{x} \in \mathbb{R}^n : \kappa||\mathbf{x}||^2 < 1\right\} \tag{1}$$

$$d^\mathbb{B} = \left(\frac{2}{1 - \kappa||\mathbf{x}||^2}\right)^2 \mathbf{I}_n \tag{2}$$

In practice, most hyperbolic neural networks are not fully hyperbolic, but approximate many of their operations in $\mathcal{T}_\mathbf{o}\mathbb{B}_\kappa^n$, i.e., in Euclidean space along a vector $\mathbf{u}$ tangential to the origin of $\mathbb{B}_\kappa^n$ (Chen et al., 2022). The corresponding exponential and logarithmic maps are defined as

$$\exp_\mathbf{o}^\kappa(\mathbf{u}) = \tanh\left(\sqrt{\kappa}||\mathbf{u}||\right) \frac{\mathbf{u}}{\sqrt{\kappa}||\mathbf{u}||} \tag{3}$$

$$\log_\mathbf{o}^\kappa(\mathbf{v}) = \frac{1}{\sqrt{\kappa}} \tanh^{-1}\left(\sqrt{\kappa}||\mathbf{v}||\right) \frac{\mathbf{v}}{||\mathbf{v}||} \tag{4}$$

For instance, $\exp_\mathbf{o}^\kappa(f(\log_\mathbf{o}^\kappa(\mathbf{x}))$ approximates a (possibly undefined) hyperbolic function $\hat{f}$ for hyperbolic embeddings $\mathbf{x}$ by solving its Euclidean equivalent f in the tangent space of the origin and projecting the result back to $\mathbb{B}_\kappa^n$.

3.2 Hyperbolic hierarchical text classification

The general idea of our architecture, shown in Figure 1, is to extract the node features from the text embeddings and use them as input to a hyperbolic graph convolutional network, which is defined as follows.

Given a label hierarchy, i.e., a tree or directed acyclic graph $G(V, E)$ with vertices V and edges E, multi-label text classification can be solved as a node classification problem using message passing between parent and child labels using a Graph Convolutional Network (Kipf and Welling, 2017). The Hyperbolic Graph Convolutional Network (HGCN) performs the neighborhood aggregation in the tangent space and uses a different trainable curvature κ in each layer (Chami et al., 2019). Finally, Hyperbolic Informed Embeddings improve the embedding quality of an L-layer HGCN by aligning the root node with the origin of the tangent space and stretching the nodes via a loss function $\mathcal{L}_{HIE}$ on the final hidden state $\mathbf{H}^{(L)}$ (Yang et al., 2023).

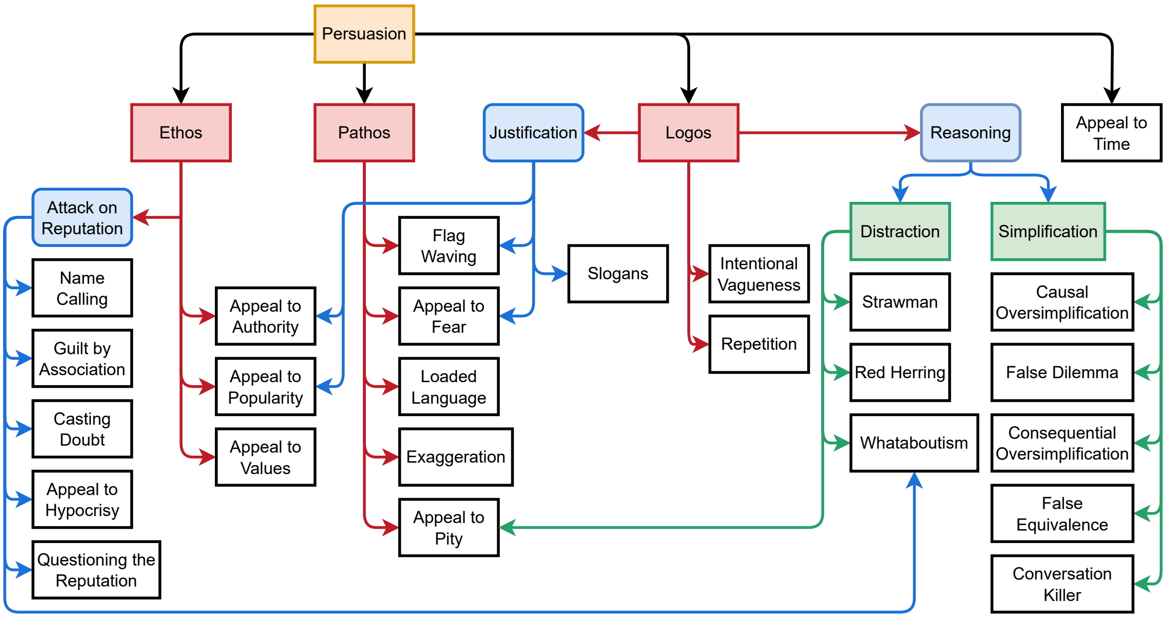

Figure 2: The hierarchy of persuasion techniques defined by SemEval 2024 (Dimitrov et al., 2024) adapted to the taxonomy defined by the Slavic NLP 2025 shared task (Piskorski et al., 2025). **Persuasion** is the root node, **Ethos, Pathos, Logos** are first level categories, **Attack on Reputation, Justification, Reasoning** are the second level categories, **Distraction, Simplification** are third level categories, and the rest are leaf nodes.

The input node features $\mathbf{H}^{(0)} \in \mathbb{R}^{|V| \times n}$ are extracted from a text embedding such as the output of XLM-RoBERTa (Conneau et al., 2020). In particular, text embeddings $\mathbf{u} \in \mathbb{R}^k$ are interpreted as points in the tangent space, projected to a vector $\tilde{\mathbf{h}}^{(0)} \in \mathbb{R}^{|V|n}$ via

$$\tilde{\mathbf{h}}^{(0)} = \mathbf{\Pi u}, \ \mathbf{\Pi} \in \mathbb{R}^{|V|n \times k} \tag{5}$$

and then reshaped into dimension $|V| \times n$.

Model outputs and gold labels are used to minimize binary cross-entropy loss $\mathcal{L}_{BCE}$, as is usual for multi-label classification tasks, and the final loss function $\mathcal{L} = \mathcal{L}_{BCE} + \mathcal{L}_{HIE}$ is optimized with Riemannian Adam, which generalizes to arbitrary Riemannian manifolds, including Euclidean and hyperbolic space (Becigneul and Ganea, 2019).

4 Experimental Setup

Although the Slavic NLP 2025 shared task defines a label taxonomy by grouping related persuasion techniques into categories such as Justification and Simplification (Piskorski et al., 2025), we base our graph G on the deeper hierarchy introduced at SemEval 2024 (Dimitrov et al., 2024) to feed the model with even more contextual information. The following additions are made to this hierarchy: Appeal to Time, sometimes called Kairos, is added as a fourth mode of persuasion next to Ethos, Pathos, and Logos. False Equivalence is a child node of Simplification. Appeal to Pity, called Appeal to Emotion in the original hierarchy, gets an additional edge from Pathos, as defined in the shared task taxonomy. Other labels are mapped to their corresponding synonyms, e.g., Bandwagon is mapped to Appeal to Popularity. The resulting graph is shown in Figure 2.

All models in the following, including a non-hierarchical baseline, use domain-adapted XLM-R-parla embeddings (Mochtak et al., 2024) pretrained on the ParlaMint 3.0 dataset (Erjavec et al., 2023) in 30 European languages. They are optimized with Riemannian Adam (Becigneul and Ganea, 2019) using 0.01 weight decay for 5 epochs. The first epoch is used for a linear warmup to a peak learning rate of $1e^{-5}$, and the remaining epochs use a cosine annealing schedule. Due to label imbalance, positive labels y are weighted with a factor $\frac{\#\text{neg}(y)}{\#\text{pos}(y)}$ during the calculation of the binary cross-entropy loss, where the negative and positive frequencies are estimated on the training split. Using unweighted loss or a learning rate of $2e^{-5}$ or greater leads to the model not learning anything. The first 50% of the language model layers are frozen to significantly decrease memory and time requirements for a minor trade-off in accuracy. After training, the checkpoint with the best *macro F1* score is loaded.

All labeled data from SemEval 2021 (Dimitrov et al., 2021), SemEval 2023 (Piskorski et al., 2023), and SemEval 2024 (Dimitrov et al., 2024) is added to the training data, covering a total of 14 languages, including Macedonian in addition to the five relevant Slavic ones (Bulgarian, Croatian, Polish, Russian, Slovene). Greek, Georgian, and Arabic texts are removed due to their different alphabets. 20% of the data are used for validation. The remaining 80% are augmented with machine translations from all languages into the six present Slavic languages and English, the language which makes up the majority of the available human-written text. The result is a large augmented training split with equal proportions in these seven languages and additional human-authored data in French, German, Italian, and Spanish. Even though the test data includes only five Slavic languages and some persuasion techniques likely get lost in translation, we expect that the additional languages and machine-translated text using Latin and Cyrillic scripts will improve the results due to the cross-lingual nature of XLM-RoBERTa. The translations are generated using the MADLAD-400-3B model (Kudugunta et al., 2023) with 4-bit quantization.

We use the PyTorch Geometric implementation of GCNs (Fey and Lenssen, 2019), and the HGCN and HIE implementations by Chami et al. (2019) and Yang et al. (2023). To some models, we append a linear layer to combine the output features of all nodes, rather than classifying directly in the GCN. All models are trained for Subtask 2 (multi-label classification). For Subtask 1 (binary detection), we simply check whether the hierarchical model predicts at least one leaf node. In addition to binary, micro, and macro F_1 scores, we compute hierarchical F_H scores, which are equivalent to the micro F_1 scores with all ancestors added to the predictions and gold labels, thus "punishing" the model less when it predicts a wrong leaf node but a correct parent label (Kosmopoulos et al., 2015).

Model	Dim	F_H	Micro	Macro
Baseline	—	—	18.22	14.59
GCN	512	27.39	13.89	11.93
HGCN	512	29.53	14.67	12.68
HGCN+L	512	31.49	16.58	13.27
HIE	256	29.96	14.82	12.79
HIE	512	29.87	14.83	12.85
HIE+L	256	**31.69**	**16.76**	13.33
HIE+L	512	31.62	16.65	**13.28**

Table 1: Hierarchical, micro, and macro F_1 scores in % on the validation set using different node dimensions, average over 3 runs. The suffix +L denotes an additional linear output layer. The best results of the proposed architecture are marked in **bold**.

5 Results

The results on the validation set for a standard non-hierarchical XLM-RoBERTa classifier baseline and a selection of 3-layer hierarchical classifiers using GCNs, HGCNs, and HGCNs with HIE loss are shown in Table 1. Experiments with the suffix +L use an additional linear output layer.

Unfortunately, neither of the trained hierarchical models meets the baseline. On the other hand, it can be seen that HGCNs perform in fact slightly better than their Euclidean counterparts, especially when HIE loss is applied. The curvature of the final layer converges to 0.7956 in most experiments. The extended SemEval 2024 label hierarchy increases the number of labels from 25 to 34 ($+36\%$), while at the same time, the micro F_1 scores approximately double when moving from leaf-only to hierarchical predictions. While the models struggle with predicting the correct fine-grained persuasion techniques, such as Causal Oversimplification, they classify the more general categories, such as Simplification or Logos, more reasonably well.

Granted, this is not a helpful functionality within the scope of the shared task. However, in a scenario where a more general label is more valuable than the absence of a prediction, such a hierarchical model provides a fallback solution by making use of granular contextual information to detect propaganda techniques on different levels of granularity.

Furthermore, the scores appear to generally improve when passing the node features after the last graph convolutional layer through a linear layer, rather than solving the problem as a pure node classification task. However, this is contested by Table 2, which shows the results on the test set split by language. Here, node classification models typically outperform those with an added linear layer. There is no obvious pattern in the choice of the node dimension, either: Although the model performs better in some languages using a dimension of 256, it performs better in other languages using a 512-dimensional Poincaré ball.

Despite the low multi-label F_1 scores, the binary F_1 scores indicate again that not the detection of persuasion techniques is the challenging part of this task, but the classification of the correct fine-grained label. Out of seven systems, the presented

Language	Model	Binary				Multi – Micro				Multi – Macro			
		P	R	F_1	Rank	P	R	F_1	Rank	P	R	F_1	Rank
Bulgarian	HIE-256	85.6	85.2	85.4	—	12.2	84.5	21.3	6	12.2	83.3	19.4	3
	HIE-512	81.1	92.7	86.5	4	11.4	90.1	20.3	—	11.2	90.3	18.9	—
Croatian	HIE-256	96.9	91.2	93.9	2	23.0	85.4	36.2	3	21.6	88.8	32.4	—
	HIE-512	82.5	97.1	89.2	—	21.0	97.5	34.6	—	21.2	96.3	32.7	2
Polish	HIE-512+L	85.5	93.3	89.2	4	14.3	86.2	24.6	6	14.0	81.0	22.5	5
Russian	HIE-256	84.8	80.6	82.7	—	6.9	75.1	12.6	—	6.7	70.6	11.2	5
	HIE-512	82.9	88.7	85.7	3	6.7	73.8	12.2	—	6.5	69.3	11.0	—
	HIE-512+L	81.6	87.8	84.6	—	7.4	73.2	13.5	5	6.5	64.4	11.1	—
Slovenian	HIE-256	82.4	89.2	85.6	1	9.8	74.6	17.4	6	9.0	84.8	14.9	3

Table 2: The best models on the test set for Subtask 1 (binary classification) and Subtask 2 (multi-label classification). We show only the models that made it into the official rankings, and present their precision, recall, F_1 scores, and ranks in the corresponding task. The model name includes the node dimension, and the suffix +L denotes an additional linear output layer. 7 teams participated.

architecture achieved an average rank of 2.8 in the binary classification task, and average ranks of 5.2 and 3.6 with respect to the micro and macro averages in the multi-label classification task.

The large class-averaged recall scores and small precision scores outline the main problem of the trained models: Whenever propaganda is detected, the models usually predict at least four persuasion techniques for the same paragraph, and in most cases, nearly all of them. Predictions on the validation data show that this is not the case for the more general labels, which violates the hierarchical property that parent labels are at least as likely to appear as their child labels. However, while removing predictions based on this criterion slightly increases precision, it significantly decreases recall and the F_1 score. Due to the difficulty of the task, it is not clear whether the model can be improved with more careful parameter tuning, or whether the task itself is not suited for this approach. The added categories are possibly too abstract and cover too different persuasion techniques to model strong hierarchical relationships.

While not explicitly shown here, early experiments indicated that the machine translated data significantly improves the model's capability to identify underrepresented labels, in particular those that have been added in more recent datasets. This is supported by the fact that our baseline model outperforms the official shared task baseline, which uses the same architecture but is trained only on the shared task data.

Finally, we experiment by weighting the loss function

$$\mathcal{L} = \mathcal{L}_{BCE} + \lambda \mathcal{L}_{HIE}$$

with a weighing factor $\lambda > 0$ and observe no difference for $\lambda \in \{0.1, 1\}$.

6 Conclusion

We have presented an architecture for hierarchical text classification based on the extraction of latent node features from text embeddings and passing messages between these nodes in a hierarchy-encoding graph convolutional network. We have further shown that making use of hyperbolic geometry improves the quality of these node embeddings. However, while our results are comparable to those of other participants in the Slavic NLP 2025 shared task, this architecture has yet to meet the baseline given by a simple non-hierarchical XLM-RoBERTa classifier trained on the same augmented data. The granular classification of propaganda remains a challenging task.

On the other hand, the hierarchy of persuasion techniques is highly abstract by nature and possibly does not model strong hierarchical relationships between the labels. Given the benefit of the hierarchical approach that the trained model can fall back to more general predictions when it fails to predict the correct granular label, the proposed architecture is still an interesting candidate for related downstream tasks with stronger baselines such as multi-label hierarchical subject classification.

Limitations

This research is a contribution to the Slavic NLP 2025 shared task. Not all technicalities of the model architecture, such as the mathematical background, can be presented in full detail in the short paper format. However, the training parameters are described in sufficient detail to reproduce the results using the code in the provided GitHub repository. Due to time and resource constraints prior to the test phase, the results are not necessarily representative of a carefully tuned model.

Acknowledgments

This research was supported by SVV project number 260 821 and the European Union's Horizon Europe research and innovation programme under grant agreement No. 101061016. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them. We thank the organizers and reviewers of the Slavic NLP 2025 workshop for their active communication and feedback.

References

Ivana Balazevic, Carl Allen, and Timothy Hospedales. 2019. Multi-relational poincaré graph embeddings. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Gary Becigneul and Octavian-Eugen Ganea. 2019. Riemannian adaptive optimization methods. In *International Conference on Learning Representations*.

Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. 2019. Hyperbolic graph convolutional neural networks. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Weize Chen, Xu Han, Yankai Lin, Hexu Zhao, Zhiyuan Liu, Peng Li, Maosong Sun, and Jie Zhou. 2022. Fully hyperbolic neural networks. *Preprint*, arXiv:2105.14686.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. *Preprint*, arXiv:1911.02116.

Giovanni Da San Martino, Alberto Barrón-Cedeño, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020. SemEval-2020 task 11: Detection of propaganda techniques in news articles. In *Proceedings of the Fourteenth Workshop on Semantic Evaluation*, pages 1377–1414, Barcelona (online). International Committee for Computational Linguistics.

Dimitar Dimitrov, Firoj Alam, Maram Hasanain, Abul Hasnat, Fabrizio Silvestri, Preslav Nakov, and Giovanni Da San Martino. 2024. SemEval-2024 task 4: Multilingual detection of persuasion techniques in memes. In *Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024)*, pages 2009–2026, Mexico City, Mexico. Association for Computational Linguistics.

Dimitar Dimitrov, Bishr Bin Ali, Shaden Shaar, Firoj Alam, Fabrizio Silvestri, Hamed Firooz, Preslav Nakov, and Giovanni Da San Martino. 2021. SemEval-2021 task 6: Detection of persuasion techniques in texts and images. In *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, pages 70–98, Online. Association for Computational Linguistics.

Tomaž Erjavec, Matyáš Kopp, Maciej Ogrodniczuk, Petya Osenova, Darja Fišer, Hannes Pirker, Tanja Wissik, Daniel Schopper, Martin Kirnbauer, Nikola Ljubešić, Peter Rupnik, Michal Mochtak, Henk van der Pol, Griet Depoorter, Kiril Simov, Vladislava Grigorova, Ilko Grigorov, Bart Jongejan, Dorte Haltrup Hansen, and 64 others. 2023. Multilingual comparable corpora of parliamentary debates ParlaMint 3.0. Slovenian language resource repository CLARIN.SI.

Matthias Fey and Jan Eric Lenssen. 2019. Fast graph representation learning with pytorch geometric. *Preprint*, arXiv:1903.02428.

Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*.

Aris Kosmopoulos, Ioannis Partalas, Eric Gaussier, Georgios Paliouras, and Ion Androutsopoulos. 2015. Evaluation Measures for Hierarchical Classification: a unified view and novel approaches. *Data Mining and Knowledge Discovery*, 29(3):820–865.

Sneha Kudugunta, Isaac Caswell, Biao Zhang, Xavier Garcia, Christopher A. Choquette-Choo, Katherine Lee, Derrick Xin, Aditya Kusupati, Romi Stella, Ankur Bapna, and Orhan Firat. 2023. Madlad-400: A multilingual and document-level large audited dataset. *Preprint*, arXiv:2309.04662.

Michal Mochtak, Peter Rupnik, and Nikola Ljubešić. 2024. The parlasent multilingual training dataset for sentiment identification in parliamentary proceedings. *Preprint*, arXiv:2309.09783.

Wei Peng, Tuomas Varanka, Abdelrahman Mostafa, Henglin Shi, and Guoying Zhao. 2022. Hyperbolic deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):10023–10044.

Jakub Piskorski, Dimitar Dimitrov, Filip Dobrani'c, Marina Ernst, Jacek Haneczok, Ivan Koychev, Ivo Moravski, Nikola Ljubeši'c, Michał Marci'nczuk, Arkadiusz Modzelewski, and Roman Yangarber. 2025. SlavicNLP 2025 Shared Task: Detection and Classification of Persuasion Techniques in Parliamentary Debates and Social Media. In *Proceedings of the 10th Workshop on Slavic Natural Language Processing 2025 (SlavicNLP 2025)*, Vienna, Austria. Association for Computational Linguistics.

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2023. SemEval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multi-lingual setup. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2343–2361, Toronto, Canada. Association for Computational Linguistics.

Menglin Yang, Min Zhou, Rex Ying, Yankai Chen, and Irwin King. 2023. Hyperbolic representation learning: Revisiting and advancing. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 39639–39659. PMLR.

Jie Zhou, Chunping Ma, Dingkun Long, Guangwei Xu, Ning Ding, Haoyu Zhang, Pengjun Xie, and Gongshen Liu. 2020. Hierarchy-aware global model for hierarchical text classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1106–1117, Online. Association for Computational Linguistics.