Maximally Expressive GNNs for Outerplanar Graphs

Franka Bause* [1,2], Fabian Jogl* [3,4], Patrick Indri[3], Tamara Drucks[3], David Penz[3,5], Nils M. Kriege[1,6], Thomas Gärtner[3], Pascal Welke[7,3], Maximilian Thiessen[3]

[1] Faculty of Computer Science, University of Vienna, Vienna, Austria
[2] UniVie Doctoral School Computer Science, University of Vienna, Vienna, Austria
[3] Machine Learning Research Unit, TU Wien, Vienna, Austria
[4] Center for Artificial Intelligence and Machine Learning, Vienna, Austria
[5] Multimedia Mining and Search, Johannes Kepler University Linz, Linz, Austria
[6] Research Network Data Science, University of Vienna, Vienna, Austria
[7] Lancaster University Leipzig, Leipzig, Germany
{firstname.lastname}@{univie, tuwien}.ac.at

Abstract. We propose a *linear time* graph transformation that enables the Weisfeiler-Leman algorithm and message passing graph neural networks to be maximally expressive on *outerplanar* graphs. Our approach is motivated by the fact that most pharmaceutical molecules correspond to outerplanar graphs. Existing research predominantly enhances the expressivity of graph neural networks without specific graph classes in mind, which often leads to methods that are impractical due to their computational complexity. In contrast, the restriction to outerplanar graphs enables to encode the Hamiltonian cycle of each biconnected component in linear time. We prove that our method achieves maximum expressivity on outerplanar graphs. Experiments confirm that our graph transformation improves the predictive performance of MPNNs on molecular benchmark datasets at negligible computational overhead.

Keywords: Graph Neural Networks · Molecular Property Prediction · Expressivity.

1 Introduction

We study graph neural networks (GNNs) for the class of outerplanar graphs and devise a linear time pre-processing step that enables message passing graph neural networks (MPNNs) to distinguish all non-isomorphic outerplanar graphs. It was shown that MPNNs have limited *expressivity*, i.e., there exist non-isomorphic graphs that each MPNN must map to identical embeddings [3]. This led to the development of GNNs that are more expressive than MPNNs, often called *higher-order* GNNs, which usually come with a significant increase in computational complexity. However, for certain domains of interest, the graph structure can

*Equal contribution.

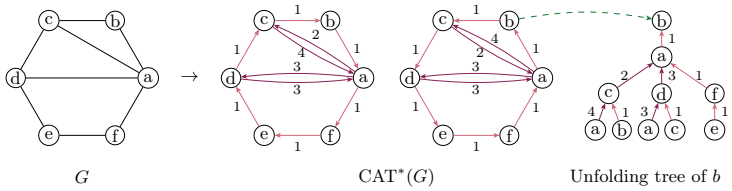

Fig. 1. CAT* transforms a biconnected outerplanar graph by copying it and directing the edges of its (unique) Hamiltonian cycle once in each direction. All other edges are present in both directions and are annotated with the distance of their endpoints on the Hamiltonian cycle. This encodes the HAL sequence (which can be used to uniquely identify the graph) in the unfolding tree of each node.

be exploited to build efficient higher-order GNNs. In this work, we focus on the pharmaceutical domain and on graphs that represent molecules. Over 92% to 97% of the graphs in widely used benchmark datasets in this domain are *outerplanar*. We focus on this class of graphs and devise a linear time transformation that allows MPNNs to become maximally expressive on outerplanar graphs. This implies that, in principle, our architecture can solve any learning task on outerplanar graphs. Our experiments show that our proposed transformation improves the predictive performance of several GNN architectures on multiple benchmark learning tasks with little increase in runtime.

2 CAT* and CAT

We build on the results of Colbourn and Booth [2] to first transform biconnected outerplanar graphs to make them distinguishable by the Weisfeiler-Leman algorithm. We call this first step cyclic adjacency transformation (CAT*). Formal definitions and proofs can be found in the full paper [1]. CAT* transforms biconnected outerplanar graphs by augmenting two copies of the graph, directing the edges of the unique Hamiltonian cycle, and annotating them with the distance of their corresponding nodes. This way, so-called Hamiltonian adjacency list (HAL) sequences are encoded in the Weisfeiler-Leman unfolding trees. For two biconnected outerplanar graphs, these sequences are the same or cyclic shifts of each other, if and only if the graphs are isomorphic [2]. Figure 1 shows an overview of CAT* and how the HAL sequences are encoded in the unfolding trees of transformed graphs. We extend this to CAT by transforming all biconnected outerplanar components of outerplanar graphs and adding additional vertices that help determine the orientation of these components in the original graph, as shown in Figure 2. Our transformation CAT enables the Weisfeiler-Leman algorithm to distinguish all non-isomorphic outerplanar graphs.

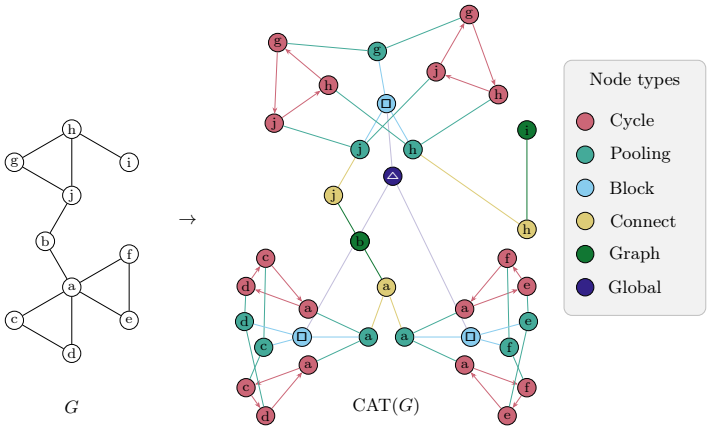

Fig. 2. CAT takes a graph, applies CAT* to its biconnected outerplanar components, and adds further nodes to encode their position and orientation within the graph. Letters represent original node ids and colors represent edge and node labels from CAT. *Cycle nodes* represent nodes added by CAT*, *pooling nodes* connect two corresponding cycle nodes, and *block nodes* connect the pooling nodes for each biconnected component, thereby able to store information about its HAL sequence. A *global node* connects these nodes to reduce the diameter of the graph. *Connect nodes* connect pooling nodes to the rest of the graph, determining the orientation of the biconnected component.

3 Experimental Evaluation

We investigate whether our proposed graph transformation CAT can improve the predictive performance of MPNNs on molecular benchmark datasets.[††] We utilize three commonly used MPNNs on ten datasets with and without CAT. Our full paper [1] presents these experiments in detail. Table 1 shows the predictive performance. Note that our baseline models obtain strong results, often surpassing the performance of (higher-order) GNNs reported in the literature, and that we train each MPNN and MPNN+CAT with exactly the same sets of hyperparameters. Overall, CAT improves the predictive performance of GIN and GCN in the majority of datasets (6/10 and 8/10, respectively). For GIN and GCN, performance increases reliably on all datasets, except `MOLLIPO` and `MOLTOX21`. Surprisingly, CAT does not work well with GAT and only improves its performance in 2/10 datasets. Most notably on `ZINC`, CAT achieves very strong results boosting the predictive performance of MPNNs between 33% for GCN and even 46% for GAT.

4 Conclusion

We propose a graph transformation that enables the Weisfeiler-Leman algorithm to be maximally expressive on outerplanar graphs. We rely on the fact that bi-

[††]Our code can be found at https://github.com/ocatias/outerplanarGNN.

Table 1. Predictive performance of MPNNs with and without CAT. Arrows indicate whether smaller (↓) or bigger (↑) results are better. **Bold** entries are an MPNN with CAT that outperforms the same MPNN without CAT.

Dataset →	ZINC	ZINC250k	MOLHIV	MOLBACE	MOLBBBP
↓ Model	MAE ↓	MAE ↓	ROC-AUC ↑	ROC-AUC ↑	ROC-AUC ↑
GIN	0.168 ± 0.007	0.033 ± 0.003	77.9 ± 1.0	74.6 ± 3.2	66.0 ± 2.1
CAT+GIN	$\mathbf{0.101 \pm 0.004}$	0.034 ± 0.003	76.7 ± 1.8	$\mathbf{79.5 \pm 2.5}$	$\mathbf{67.2 \pm 1.8}$
GCN	0.184 ± 0.013	0.067 ± 0.005	76.7 ± 1.4	77.9 ± 1.7	66.1 ± 2.4
CAT+GCN	$\mathbf{0.123 \pm 0.008}$	$\mathbf{0.034 \pm 0.003}$	$\mathbf{77.1 \pm 1.6}$	$\mathbf{79.2 \pm 1.5}$	$\mathbf{68.3 \pm 1.7}$
GAT	0.375 ± 0.013	0.103 ± 0.004	76.6 ± 2.0	81.7 ± 2.3	66.2 ± 1.4
CAT+GAT	$\mathbf{0.201 \pm 0.022}$	$\mathbf{0.046 \pm 0.004}$	75.3 ± 1.6	79.3 ± 1.6	66.0 ± 1.9
Dataset →	MOLSIDER	MOLESOL	MOLTOXCAST	MOLLIPO	MOLTOX21
↓ Model	ROC-AUC ↑	RMSE ↓	ROC-AUC ↑	RMSE ↓	ROC-AUC ↑
GIN	56.6 ± 1.0	1.105 ± 0.077	65.3 ± 0.6	0.717 ± 0.016	75.8 ± 0.7
CAT+GIN	$\mathbf{58.2 \pm 0.9}$	$\mathbf{0.985 \pm 0.055}$	$\mathbf{65.6 \pm 0.5}$	0.798 ± 0.031	74.8 ± 1.0
GCN	56.7 ± 1.5	1.053 ± 0.087	64.4 ± 0.4	0.748 ± 0.018	76.4 ± 0.3
CAT+GCN	$\mathbf{57.9 \pm 1.8}$	$\mathbf{1.006 \pm 0.036}$	$\mathbf{66.2 \pm 0.8}$	0.771 ± 0.023	74.9 ± 0.8
GAT	58.4 ± 1.0	1.037 ± 0.063	63.8 ± 0.8	0.728 ± 0.024	76.3 ± 0.6
CAT+GAT	58.3 ± 1.3	1.09 ± 0.048	$\mathbf{64.5 \pm 0.8}$	0.754 ± 0.021	75.4 ± 0.7

connected outerplanar graphs can be uniquely identified by their Hamiltonian adjacency list sequences, which CAT encodes in unfolding trees. Our work highlights the value of GNNs designed for specific graph classes. In general, achieving high expressivity is computationally expensive. We have demonstrated that for *outerplanar* graphs maximal expressivity can be achieved in linear time.

Acknowledgments. This work was supported in part by the Vienna Science and Technology Fund (WWTF) through the Vienna Research Group Algorithmic Data Science for Computational Drug Discovery [10.47379/VRG19009] and the project StruDL [10.47379/ICT22059]; by the TU Wien DK SecInt; and by the Austrian Science Fund (FWF) through project NanOX-ML (6728). MT acknowledges support from a DOC fellowship of the Austrian academy of sciences (ÖAW).

References

1. Bause, F., Jogl, F., Indri, P., Drucks, T., Penz, D., Kriege, N.M., Gärtner, T., Welke, P., Thiessen, M.: Maximally expressive GNNs for outerplanar graphs. Transactions on Machine Learning Research (2025)
2. Colbourn, C.J., Booth, K.S.: Linear time automorphism algorithms for trees, interval graphs, and planar graphs. SIAM Journal on Computing **10**(1), 203–225 (1981)
3. Morris, C., Ritzert, M., Fey, M., Hamilton, W.L., Lenssen, J.E., Rattan, G., Grohe, M.: Weisfeiler and Leman go neural: Higher-order graph neural networks. In: AAAI (2019)