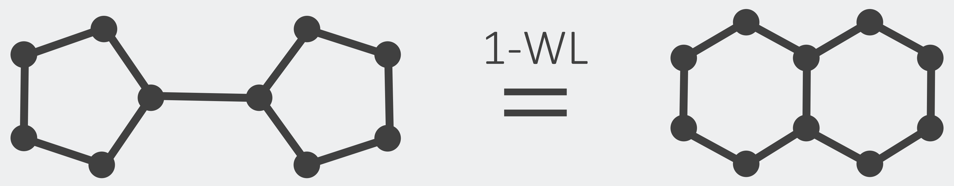
MAXIMALLY EXPRESSIVE GNNS FOR OUTERPLANAR GRAPHS

Franka Bause, Fabian Jogl, Patrick Indri, Tamara Drucks, David Penz,
Nils Kriege, Thomas Gärtner, Pascal Welke & Maximilian Thiessen

TECHNISCHE UNIVERSITÄT WIEN
universität wien

We propose a linear time graph transformation that enables the Weisfeiler-Leman test (WL) and message passing graph neural networks (MPNNs) to be maximally expressive on outerplanar graphs. Most pharmaceutical molecules correspond to outerplanar graphs. However, there are non-isomorphic outerplanar graphs, that cannot be distinguished by 1-WL. Our method relies on encoding the Hamiltonian cycle of each biconnected component and achieves maximum expressivity on outerplanar graphs.



TL;DR

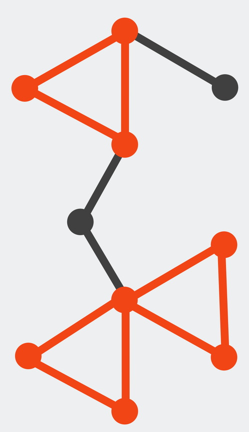
We propose a linear time preprocessing that allows MPNNs to distinguish all outerplanar graphs.

Our approach boosts predictive performance of MPNNs on a variety of molecular benchmark datasets (see table below). The transformation is linear time, i.e., no additional computational complexity is added to the training process or inference.

Dataset →	ZINC	MOLHIV	MOLBACE	MOLBBBP	MOLSIDER
↓ Model	MAE ↓	ROC-AUC ↑	ROC-AUC ↑	ROC-AUC ↑	ROC-AUC ↑
GIN	0.168 ± 0.007	77.9 ± 1.0	74.6 ± 3.2	66.0 ± 2.1	56.6 ± 1.0
CAT+GIN	0.101 ± 0.004	76.7 ± 1.8	79.5 ± 2.5	67.2 ± 1.8	58.2 ± 0.9
GCN	0.184 ± 0.013	76.7 ± 1.4	77.9 ± 1.7	66.1 ± 2.4	56.7 ± 1.5
CAT+GCN	0.123 ± 0.008	77.1 ± 1.6	79.2 ± 1.5	68.3 ± 1.7	57.9 ± 1.8
GAT	0.375 ± 0.013	76.6 ± 2.0	81.7 ± 2.3	66.2 ± 1.4	58.4 ± 1.0
CAT+GAT	0.201 ± 0.022	75.3 ± 1.6	79.3 ± 1.6	66.0 ± 1.9	58.3 ± 1.3

Dataset →	MOLESOL	MOLTOXCAST	MOLLIPO	MOLTOX21
↓ Model	RMSE ↓	ROC-AUC ↑	RMSE ↓	ROC-AUC ↑
GIN	1.105 ± 0.077	65.3 ± 0.6	0.717 ± 0.016	75.8 ± 0.7
CAT+GIN	0.985 ± 0.055	65.6 ± 0.5	0.798 ± 0.031	74.8 ± 1.0
GCN	1.053 ± 0.087	64.4 ± 0.4	0.748 ± 0.018	76.4 ± 0.3
CAT+GCN	1.006 ± 0.036	66.2 ± 0.8	0.771 ± 0.023	74.9 ± 0.8
GAT	1.037 ± 0.063	63.8 ± 0.8	0.728 ± 0.024	76.3 ± 0.6
CAT+GAT	1.090 ± 0.048	64.5 ± 0.8	0.754 ± 0.021	75.4 ± 0.7

A graph is outerplanar if it can be drawn in the plane without edge crossings and with all nodes belonging to the exterior face.



graph G

CAT* TRANSFORMATION

(simplified) CAT* transforms a biconnected outerplanar graph G into a modified graph $G' = \mathrm{CAT}^*(G)$. This is done by computing the (directed) Hamiltonian cycle C of G and its reverse, creating two directed components in G'. Edges of G not included in those are added to both components with both directions each. Edges are annotated so that the HAL sequences can be recovered from the unfolding trees.

THEOREM 1

Two biconnected outerplanar graphs G and H are isomorphic, if and only if

$$\mathrm{WL}(\mathrm{CAT}^*(G)) = \mathrm{WL}(\mathrm{CAT}^*(H))$$

THEOREM 2

Outerplanar graphs G and H are isomorphic, if and only if

$$\mathrm{WL}(\mathrm{CAT}(G)) = \mathrm{WL}(\mathrm{CAT}(H))$$

⊥ Nodes of G not in blocks
★ Node connecting node pairs of each C and its reverse
⋈ Nodes of G connecting multiple blocks
□ Node connecting all ★ nodes of a block
△ Node connecting all □ nodes

CAT TRANSFORMATION

(simplified) We define the CAT transformation by applying CAT* to the blocks of the graph G and adding further nodes and edges. The additional nodes and edges allow to recover the orientation and location of the blocks, making non-isomorphic outerplanar graphs distinguishable by 1-WL.

$\mathrm{CAT}(G)$
transformed graph G