Weisfeiler and Lehman Go Paths: Learning Topological Features via Path Complexes

Quang Truong and Peter Chin

Thayer School of Engineering, Dartmouth College

MOTIVATION

Prior works on topological higher-order GNNs often depend on assumptions about sub-structures of graphs, such as cliques, cycles, and rings.

Our study presents a novel perspective by focusing on **simple paths**, a universal element in graphs, during the topological message-passing process.

LIFTING TRANSFORMATIONS

Given a simple graph $G = (\mathcal{V}, \mathcal{E})$ with a finite vertex set $\mathcal{V}$ and edge set $\mathcal{E}$, we can apply lifting transformation such as clique [2] cell [3], or path complex lifting.

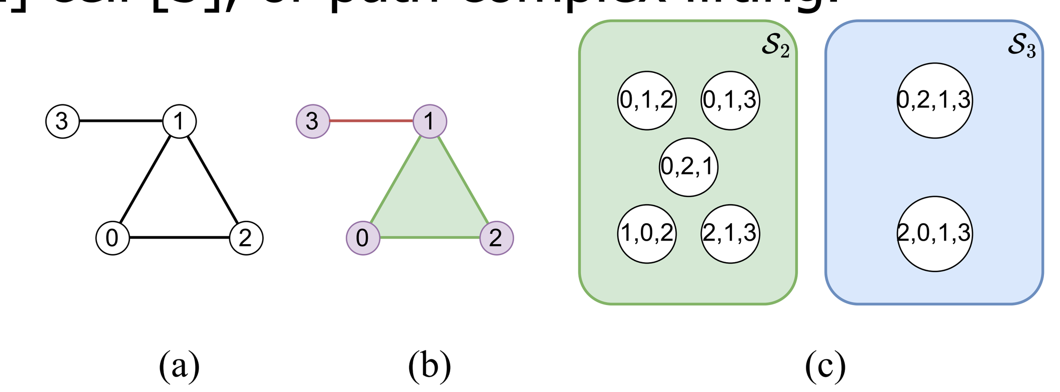

Figure 1. (a) Original graph; (b) Simplicial complex, which contains a 2-simplex, 4 1-simplices, and 4 0-simplices, arising from the original graph. Regular cell complex coincides with the simplicial complex in this case; (c) Simple path spaces $\mathcal{S}_2$ and $\mathcal{S}_3$ corresponding to the path complex arising from the original graph. Elementary paths of $\mathcal{S}_0$ and $\mathcal{S}_1$ are indeed 0-simplices (0-cells) and 1-simplices (1-cells) of the simplicial complex (regular cell complex).

TUDATASETS

Dataset	PROTEINS	NCI1	NCI109	IMDB-B
RWK	59.6 ± 0.1	> 3 days	N/A	N/A
GK (k=3)	71.4 ± 0.3	62.5 ± 0.3	62.4 ± 0.3	N/A
PK	73.7 ± 0.7	82.5 ± 0.5	N/A	N/A
WL Kernel	75.0 ± 3.1	86.0 ± 1.8 ♦	N/A	73.8 ± 3.9
DCNN	61.3 ± 1.6	56.6 ± 1.0	N/A	49.1 ± 1.4
DGCNN	75.5 ± 0.9	74.4 ± 0.5	N/A	70.0 ± 0.9
IGN	76.6 ± 5.5	74.3 ± 2.7	72.8 ± 1.5	72.0 ± 5.5
GIN	76.2 ± 2.8	82.7 ± 1.7	N/A	75.1 ± 5.1
PPGNs	77.2 ± 4.7	83.2 ± 1.1	82.2 ± 1.4	73.0 ± 5.8
Natural GN	71.7 ± 1.0	82.4 ± 1.3	N/A	73.5 ± 2.0
GSN	76.6 ± 5.0	83.5 ± 2.0	N/A	77.8 ± 3.3 ♦
pathGCN	80.4 ± 4.2 ▲	83.3 ± 1.3	N/A	N/A
PathNN	75.2 ± 3.9	82.3 ± 1.9	N/A	72.6 ± 3.3
SIN †	76.4 ± 3.3	82.7 ± 2.1	N/A	75.6 ± 3.2 ●
CIN †	77.0 ± 4.3	83.6 ± 1.4	84.0 ± 1.6 ●	75.6 ± 3.7
CAN	78.2 ± 2.0	84.5 ± 1.6	83.6 ± 1.2	N/A
CIN++	80.5 ± 3.9 ♦	85.3 ± 1.2 ▲	84.5 ± 2.4 ♦	N/A
PIN (Ours)	78.8 ± 4.4 ●	85.1 ± 1.5 ●	84.0 ± 1.5 ▲	76.6 ± 2.9 ▲

SR GRAPHS

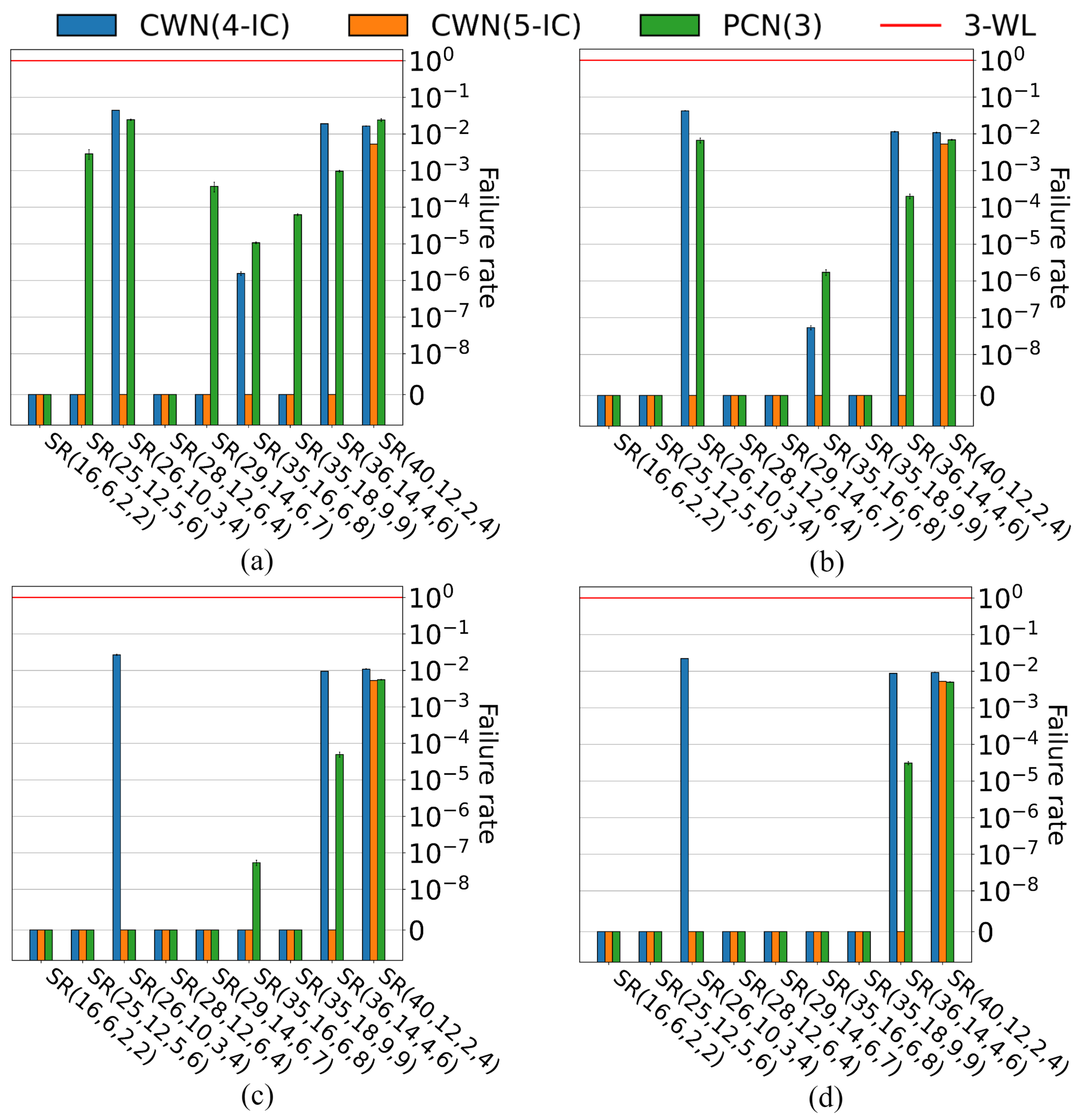

Figure 3. Failure rate comparison between CWN and PCN on SRG Families over 10 different seeds. (a) 3 MP layers. (b) 4 MP layers. (c) 5 MP layers. (d) 6 MP layers.

PATH COMPLEXES

Definition 1. (Grigor'yan et al. [1]) Given a finite non-empty set V whose element is called vertex, an **elementary p-path** on set V is any sequence of vertices with length $p+1$. Elementary p-path is denoted by $e_{i_0 \ldots i_p}$.

Definition 2. (Grigory'an et al. [1]) Given a finite non-empty set V, a **path complex** P is a non-empty collection of elementary paths such that for any sequence of vertices that belong to P, the truncated sequences, in which either the first vertex or the last vertex is removed, are also included in P.

THEORETICAL RESULTS

Proposition 3. For any two path complexes X and Y, if $\sigma \in X$ and $\tau \in Y$ have different boundary sizes $|\mathcal{B}(\sigma)| \neq |\mathcal{B}(\tau)|$, their colorings are different $c_\sigma^{X,t} \neq c_\tau^{Y,t}$ for $t > 0$.

Theorem 4. PWL is at least as powerful as SWL [2] at distinguishing non-isomorphic graphs.

Theorem 5. PWL is at least as powerful as CWL(k-IC) [3] at distinguishing non-isomorphic graphs.

Corollary 6. PWL is strictly more powerful than WL at distinguishing non-isomorphic graphs.

Corollary 7. PWL is not less powerful than 3-WL at distinguishing non-isomorphic graphs.

ILLUSTRATION

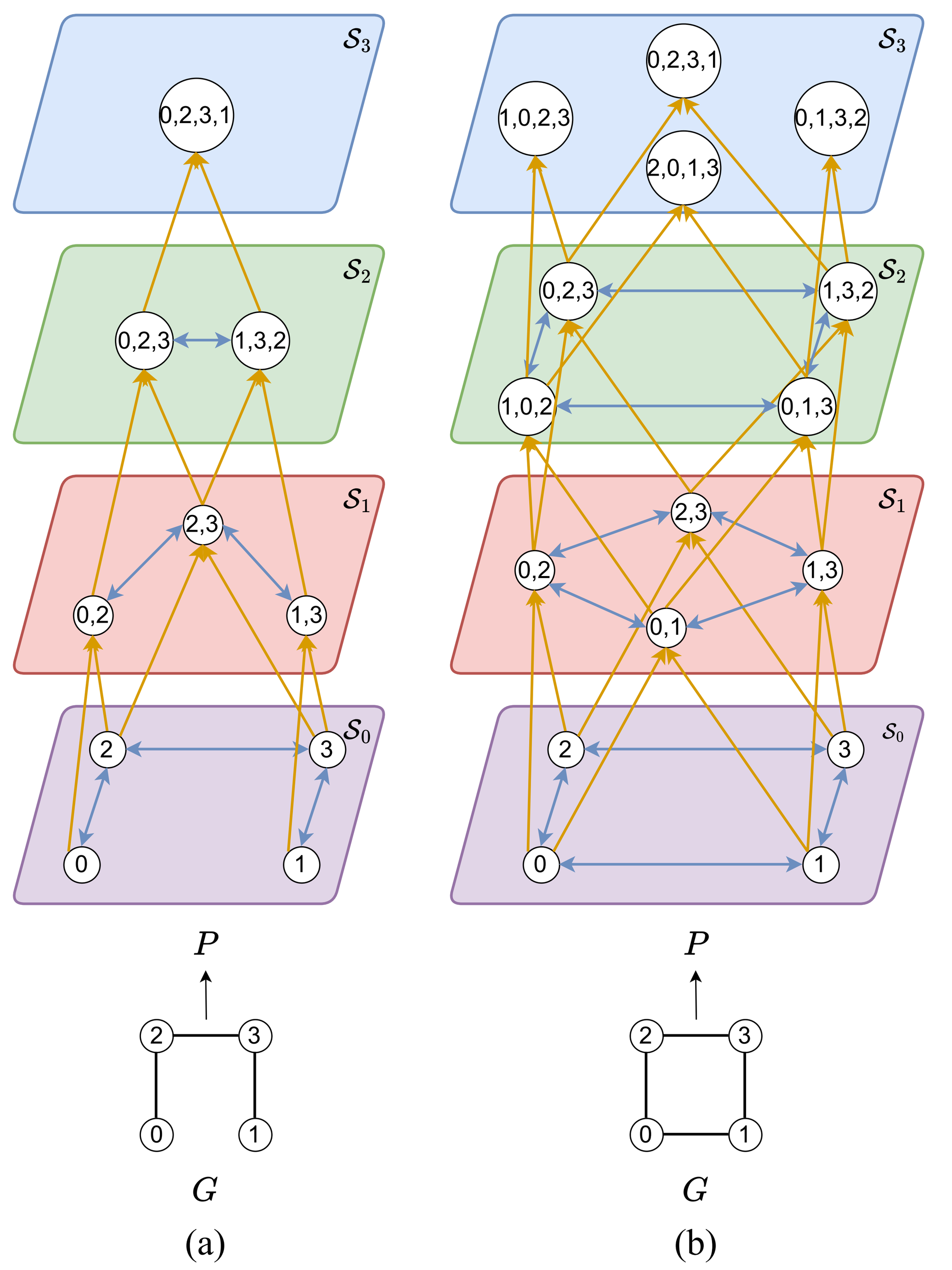

Figure 2. Examples of path complexes arising from graphs.

ZINC & OGBG-MOLHIV

Dataset	ZINC		OGBG-MOLHIV	
	No Edge Feat.	W/ Edge Feat.	Test ROC-AUC	Val. ROC-AUC
GCN	0.469 ± 0.002	N/A	N/A	N/A
GAT	0.463 ± 0.002	N/A	N/A	N/A
GatedGCN	0.422 ± 0.006	0.363 ± 0.009	N/A	N/A
GIN	0.408 ± 0.008	0.252 ± 0.014	77.07 ± 1.49	84.79 ± 0.68
PNA	0.320 ± 0.032	0.188 ± 0.004	79.05 ± 1.32	85.19 ± 0.99
DGN	0.219 ± 0.010	0.168 ± 0.003	79.70 ± 0.97	84.70 ± 0.47
HIMP	N/A	0.151 ± 0.006	78.80 ± 0.82	N/A
GSN	0.140 ± 0.006	0.115 ± 0.012	77.99 ± 1.00	**86.58 ± 0.84**
PathNN	N/A	0.090 ± 0.004	79.17 ± 1.09	N/A
CIN †	**0.115 ± 0.003**	0.079 ± 0.006	**80.94 ± 0.57**	N/A
CIN++	N/A	**0.077 ± 0.004**	80.63 ± 0.94	N/A
PIN (Ours)	0.139 ± 0.004	0.096 ± 0.006	79.44 ± 1.40	82.41 ± 0.96

REFERENCES

[1] Grigor'yan, A.; Lin, Y.; Muranov, Y.; and Yau, S.-T. 2013. Homologies of path complexes and digraphs. arXiv: 1207.2834 [math].

[2] Bodnar, C.; Frasca, F.; Otter, N.; Wang, Y.; Liò, P.; and Bronstein, M. 2021b. Weisfeiler and Lehman Go Topological: Message Passing Simplicial Networks. In Meila, M.; and Zhang, T., eds., Proceedings of the 38th International Conference on Machine Learning, volume 139 of Proceedings of Machine Learning Research, 1026–1037. PMLR.

[3] Bodnar, C.; Frasca, F.; Otter, N.; Wang, Y.; Liò, P.; Montufar, G. F.; and Bronstein, M. 2021a. Weisfeiler and Lehman Go Cellular: CW Networks. In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., Advances in Neural Information Processing Systems, volume 34, 2625-2640. Curran Associates, Inc.