EXPRESSIVENESS STUDY

A path from a source node v to a target node u is denoted by $p_{uv} = [v_1, v_2, \ldots, v_k]$, where $v_1 = v$, $v_k = u$, and $(v_i, v_{i+1}) \in E$ for $i \in \{1, \ldots, k-1\}$. Paths contain distinct vertices, and the length of the path is given by $|p_{uv}| = k - 1$, defined as the number of edges it contains. In this work, we consider paths that adhere to these criteria.

In practice, we only consider paths up to a fixed length L. Let $\mathcal{P}_{uv}^L$ denote the set of the sampled top-N optimal paths from u to v, selected based on the best path performance, with lengths not exceeding L. Recall that S and T represent the set of node pairs with observed path performance and all possible node pairs, respectively. Define $\mathcal{SP} = \bigcup_{<u,v>\in S} \mathcal{P}_{uv}^L$ as the collection of all sampled paths, and let $\mathcal{AP}$ denote the collection of all paths between the node pair combinations in T. We have $\mathcal{P}_{uv}^L \subset \mathcal{SP} \subseteq \mathcal{AP}$, where $\mathcal{SP} \to \mathcal{AP}$ as $N \to \infty$ and $S \to T$.

Definition 1 (WL-Tree rooted at v)**.** Let $G = (V, E)$. A WL-Tree W_v^L is a tree rooted at node $v \in V$ encoding the structural information captured by the 1-WL algorithm up to L iterations. At each iteration, the children of a node u are its direct neighbors, $\mathcal{N}(u) = \{w \mid (u, w) \in E\}$.

Definition 2 (Path-Tree rooted at v)**.** Let $G = (V, E)$. A Path-Tree P_v^L rooted at a node $v \in V$ is a tree of height L, where the node set at level k is the multiset of nodes that appear at position k in the paths of $\mathcal{P}_v^L$, i.e., $\{u \mid p^L(k) = u \text{ for } p^L \in \mathcal{P}_v^L\}$. Nodes at level k and level $k+1$ are connected if and only if they occur in adjacent positions k and $k+1$ in any path $p^L \in \mathcal{P}_v^L$.

Theorem 3 (DeepNT-$\mathcal{AP}$ Expressiveness Beyond 1-WL)**.** *Let $G = (V, E)$, W_v^L and W_u^L denote the WL-Trees of height L rooted at nodes $v, u \in V$, respectively. Let $f_{DeepNT}(v)$ and $f_{DeepNT}(u)$ represent the embeddings produced by DeepNT when it has access to the complete set of paths $\mathcal{AP}$. Then the following holds:*

1. *If $W_v^L \neq W_u^L$, then $f_{DeepNT}(v) \neq f_{DeepNT}(u)$.*
2. *If $W_v^L = W_u^L$, it is still possible that $f_{DeepNT}(v) \neq f_{DeepNT}(u)$.*

Proof. To prove this statement, we refer to Theorem 3.3 from Michel et al. (2023), which states that if W_v^L is structurally different from W_u^L (i.e., not isomorphic), then P_v^L is also structurally different from P_u^L. Moreover, P_v^L and P_u^L can still differ even if W_v^L and W_u^L are identical.

We first address the case where $W_v^L \neq W_u^L$. It follows that P_v^L and P_u^L will not be isomorphic. The path aggregation layer in DeepNT is straightforward to prove as injective, as it employs a permutation-invariant readout function. Consequently, DeepNT aggregates path-centric structural information from P_v^L and P_u^L to produce embeddings $f_{\text{DeepNT}}(v)$ and $f_{\text{DeepNT}}(u)$, which are guaranteed to be distinct.

Now consider the case where $W_v^L = W_u^L$. This implies that 1-WL cannot distinguish between v and u. However, P_v^L and P_u^L may still differ. The path aggregation layer of DeepNT captures path-centric structural information from P_v^L and P_u^L, resulting in distinct embeddings $f_{\text{DeepNT}}(v)$ and $f_{\text{DeepNT}}(u)$. Thus, DeepNT surpasses the expressiveness of 1-WL. This completes the proof. □

Theorem 4 (DeepNT-$\mathcal{AP}$ Distinguishes Node Pairs Beyond 1-WL)**.** *Let $G = (V, E)$, and let $\langle u, v\rangle$ and $\langle u', v'\rangle$ be two node pairs in $V \times V$ such that the local neighborhoods of u and u' are identical up to L hops, and similarly for v and v'. If the sets of paths $\mathcal{P}_{uv}^L$ and $\mathcal{P}_{u'v'}^L$ are different, DeepNT will assign distinct embeddings $f_{DeepNT}(u, v) \neq f_{DeepNT}(u', v')$.*

Theorem 4 can be readily proved by noting that the node pairs $\langle u, v\rangle$ and $\langle u', v'\rangle$ are represented by concatenated node embeddings. Since $\mathcal{P}_{uv}^L \neq \mathcal{P}_{u'v'}^L$, the distinctiveness of these representations is ensured by the expressive power of DeepNT, which surpasses that of 1-WL.

Theorem 5 (Convergence of DeepNT Predictions to True Pairwise Metrics)**.** *Let $G = (V, E)$. Suppose DeepNT is trained with an increasing number of observed node pairs $S \to T$, where $S \subseteq T = V \times V$, and the number of sampled paths $N \to \infty$. Then, the predicted pairwise metrics $\hat{y}_{uv} = f_{DeepNT}(u, v; \theta, \tilde{A})$ converge in expectation to the true metrics y_{uv}, i.e.,*

$$\lim_{S\to T} \lim_{N\to\infty} \mathbb{E}_{\langle u,v\rangle\sim T} \left[|\hat{y}_{uv} - y_{uv}|\right] = 0,$$

Proof. As $N \to \infty$, the sampled paths $\mathcal{SP}$ converge to the complete set of paths for the observed node pairs S. As $S \to T$, the set of observed node pairs expands to include all possible node pairs in $T = V \times V$. Therefore, $\mathcal{SP} \to \mathcal{AP}$ as $N \to \infty$ and $S \to T$, which means DeepNT-$\mathcal{SP}$ converges to DeepNT-$\mathcal{AP}$.

By Theorem 3 (DeepNT-$\mathcal{AP}$ Expressiveness Beyond 1-WL) and Theorem 4 (DeepNT-$\mathcal{AP}$ Distinguishes Node Pairs Beyond 1-WL), DeepNT-$\mathcal{AP}$ can uniquely identify and distinguish all node pairs based on differences in their path sets. Moreover, by Theorem 4.1 (Convergence of DeepNT's Optimization), the optimization of DeepNT converges to a stationary point $(\theta^*, \tilde{A}^*)$. This ensures that the empirical loss over the observed pairs S is minimized:

$$\mathcal{L}(\theta^*, \tilde{A}^*) = \sum_{\langle u,v \rangle \in S} l(f_{\text{DeepNT}}(u, v; \theta^*, \tilde{A}^*), y_{uv}),$$

where $l(\cdot, \cdot)$ measures the error between the predicted metrics $\hat{y}_{uv}$ and the true metrics y_{uv}. Consequently, as $S \to T$, the training error diminishes:

$$\mathbb{E}_{\langle u,v \rangle \sim S}[|\hat{y}_{uv} - y_{uv}|] \to 0.$$

As $S \to T$, the observed set S becomes dense, covering all node pairs in $T = V \times V$. Therefore, the unobserved set $T \setminus S$ becomes empty, i.e., $T \setminus S \to \emptyset$. Combined with the expressiveness of DeepNT and the completeness of path sampling, the model generalizes well to unobserved pairs $\langle u, v \rangle \in T \setminus S$. This ensures that the generalization error also diminishes:

$$\mathbb{E}_{\langle u,v \rangle \sim T \setminus S}[|\hat{y}_{uv} - y_{uv}|] \to 0.$$

Combining these results, the total error for all node pairs in T, which is the sum of the training error and the generalization error, converges to zero:

$$\epsilon_{\text{total}} = \mathbb{E}_{\langle u,v \rangle \sim S}[|\hat{y}_{uv} - y_{uv}|] + \mathbb{E}_{\langle u,v \rangle \sim T \setminus S}[|\hat{y}_{uv} - y_{uv}|] \to 0.$$

Finally, as $N \to \infty$ and $S \to T$, DeepNT-$\mathcal{SP}$ converges to DeepNT-$\mathcal{AP}$, and the predicted metrics $\hat{y}_{uv}$ converge in expectation to the true metrics y_{uv}:

$$\lim_{S \to T} \lim_{N \to \infty} \mathbb{E}_{\langle u,v \rangle \sim T} \left[|\hat{y}_{uv} - y_{uv}|\right] = 0.$$

This completes the proof. □

References

Gaspard Michel, Giannis Nikolentzos, Johannes F Lutzeyer, and Michalis Vazirgiannis. Path neural networks: Expressive and accurate graph neural networks. In *International Conference on Machine Learning*, pp. 24737–24755. PMLR, 2023.