WILTing Trees: Interpreting the Distance Between MPNN Embeddings

Masahiro Negishi[1], Thomas Gärtner[2], and Pascal Welke[2,3]

1. University of Tokyo, Japan
2. TU Wien, Austria
3. Lancaster University Leipzig, Germany

Message passing graph neural networks (MPNNs) have been reported to achieve high predictive performance in various domains [1]. To understand these performance gains, many studies have focused on the expressive power of MPNNs [2]. However, the binary nature of expressive power excludes any analysis of the distance between graph embeddings, which is considered to be a key to the predictive power of MPNNs [3]. Recently, there has been growing interest in the analysis of MPNN performance using *structural* distances between graphs that consider graph topology but ignore the target function to be learned. In contrast, we investigate the distance d_{MPNN} implicitly obtained from an MPNN and its relation to a *functional* distance d_{func} defined on the target values of the learning task.

We find that even if an MPNN was trained with classical cross-entropy loss, d_{MPNN} respects the task-relevant functional distance d_{func} and the alignment between both is highly correlated with the predictive performance of MPNNs. Since MPNNs consider graphs as multisets of Weisfeiler Leman (WL) subgraphs, we propose a method to identify WL subgraphs whose presence in a graph significantly affects its relative position to other graphs in the MPNN embedding space. Specifically, we distill MPNNs into a *weighted Weisfeiler Leman Labeling Tree* (WILT) while preserving d_{MPNN}. The WILT yields an optimal transport distance on a tree ground metric, which we prove to be a trainable generalization of the graph distances of existing high-performance graph kernels [4,5]. Figure 1 gives a brief overview of the involved concepts. We show experimentally that the WILTing tree distance fits MPNN distances well and that only a small number of WL subgraphs determine d_{MPNN}. In a qualitative experiment, the subgraphs that strongly influence d_{MPNN} are those that are known to be functionally important by domain knowledge. In short, our contributions are:

! We show that MPNN distances after training are aligned with the task-relevant functional distance of the graphs and that this is key to the high predictive performance of MPNNs.

! We propose a trainable graph distance on a weighted Weisfeiler-Lehman Labeling Tree (WILT) that generalizes Weisfeiler Leman-based distances and is efficiently computable.

! WILTs allow a straightforward definition of *relevant* subgraphs. Thus, distilling an MPNN into a WILT enables us to identify subgraphs that strongly influence the distance between MPNN embeddings, allowing an interpretation of the MPNN embedding space.

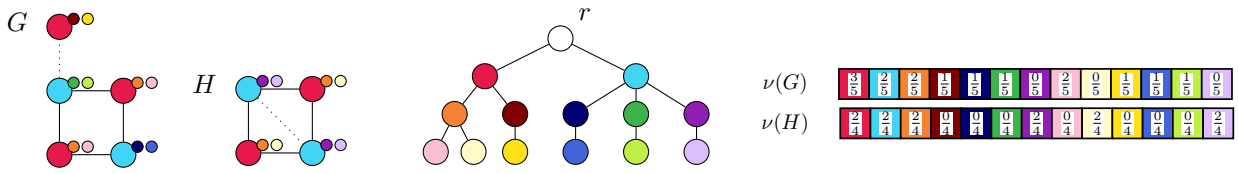

Figure 1: (left) Example of how the Weisfeiler Leman (WL) algorithm works on graphs G and H. ● and ● are colors corresponding to initial node labels. Node colors in WL-iterations one and two are shown in the small circles next to the nodes. (center): The Weisfeiler Leman Labelling Tree (WILT) built from $\mathcal{D} = \{G, H\}$, it encodes the hierarchy of WL labels and allows edge weights. (right): The WILT embeddings ν contain normalized counts of all WL colors for G and H. The optimal transport distance $d_{\mathsf{WILT}}(G, H)$ can be computed as weighted Manhattan distance of $\nu(G)$ and $\nu(H)$.

1 J. Zhou et al. *Graph neural networks: A review of methods and applications*. AI open, 1:57–81, 2020

2 K. Xuet al. *How powerful are graph neural networks?* ICLR, 2019.

3 C. Morris et al. *Position: Future directions in the theory of graph machine learning*. ICML, 2024.

4 N. Kriege et al. *On valid optimal assignment kernels and applications to graph classification*. NeurIPS, 2016.

5 M. Togninalli et al. *Wasserstein Weisfeiler-Lehman graph kernels*. NeurIPS, 2019.