A New Perspective on "How Graph Neural Networks Go Beyond Weisfeiler-Lehman?"

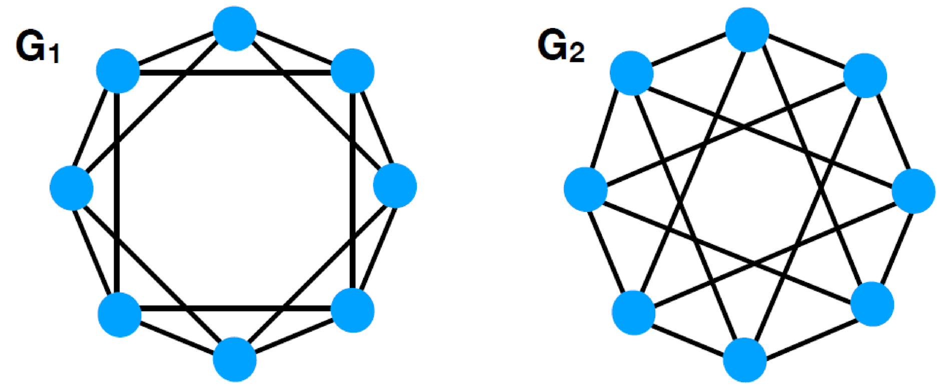
Asiri Wijesinghe and Qing Wang

School of Computing, The Australian National University, Canberra ACT 0200, Australia

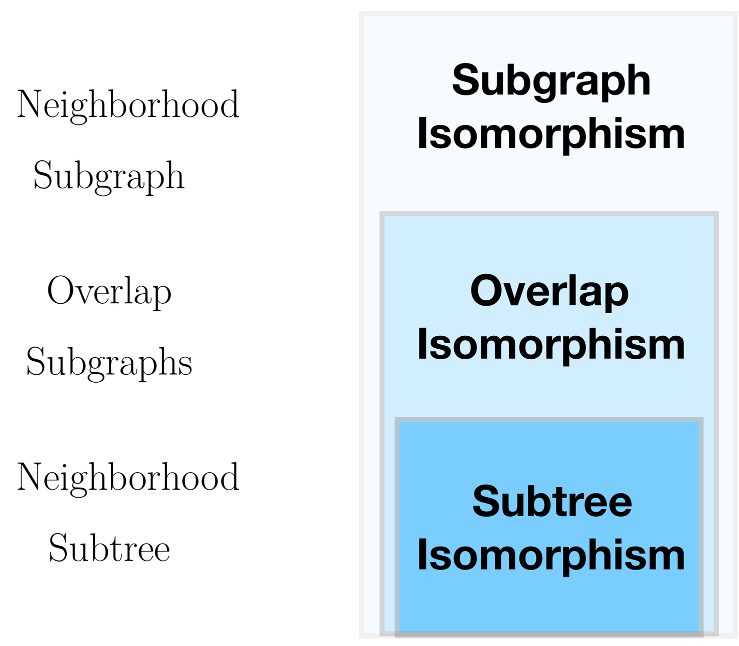
{asiri.wijesinghe, qing.wang}@anu.edu.au

Introduction

- How to design expressive yet simple GNNs that can go beyond the WL test with a theoretically provable guarantee?



A New Hierarchy of Local Isomorphism



Theorem

If $S_i \simeq_{subgraph} S_j$, then $S_i \simeq_{overlap} S_j$, but not vice versa.

If $S_i \simeq_{overlap} S_j$, then $S_i \simeq_{subtree} S_j$, but not vice versa.

Structural Coefficients

- For each vertex v and its neighbors u, we define *structural coefficients* $A_{vu} = \omega(S_v, S_{vu})$ satisfying three desirable properties:

(a) Local closeness (b) Local denseness (c) Isomorphic invariant

- An instance:

$$A_{vu} = \frac{|E_{vu}|}{|V_{vu}| \cdot |V_{vu} - 1|}|V_{vu}|^{\lambda},\ \lambda > 0$$

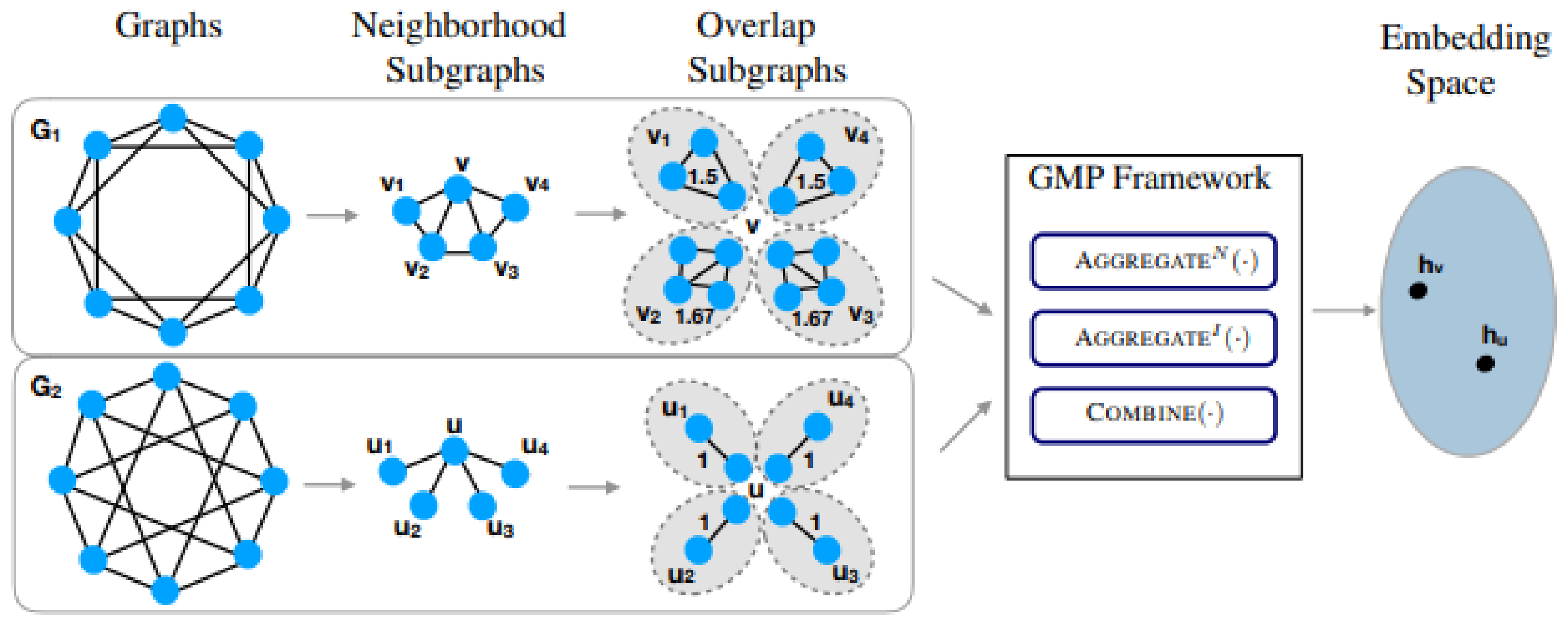
GraphSNN - A GNN Model Beyond 1-WL

- A single layer:

$$h_v^{(t)} = \textsc{Mlp}\Big(\gamma^{(t)}\Big(\sum_{u\in\mathcal{N}(v)} \tilde{A}_{vu} + 1\Big)h_v^{(t-1)} + \sum_{u\in\mathcal{N}(v)} \Big(\tilde{A}_{vu} + 1\Big)h_u^{(t-1)}\Big)$$

- Multiple layers (same as GIN):

$$h_G = \textsc{Concat}(\textsc{Readout}(\{\{h_v^{(t)} | v \in V\}\}) | t = 1, \dots, k)$$



A Generalised Message Passing GNN

- Aggregate 'messages' from neighbors $\mathcal{N}(v)$

$$\cancel{h^{(t)} = \textsc{Aggregate}\Big(\{h_u^{(t)} | u \in \mathcal{N}(v)\}\Big)}$$

$$\hookrightarrow m_a^{(t)} = \textsc{Aggregate}^N\Big(\{(\tilde{A}_{vu}, h_u^{(t)}) | u \in \mathcal{N}(v)\}\Big)$$

$$\hookrightarrow m_v^{(t)} = \textsc{Aggregate}^I\Big(\{\tilde{A}_{vu} | u \in \mathcal{N}(v)\}\Big)h_v^{(t)}$$

- Combine with its own 'message' $h_v^{(t)}$

$$\cancel{h_v^{(t+1)} = \textsc{Combine}\Big(h_v^{(t)}, h^{(t)}\Big)}$$

$$\hookrightarrow h_v^{(t+1)} = \textsc{Combine}\Big(m_v^{(t)}, m_a^{(t)}\Big)$$

Numerical Experiments

- Classification on Open Graph Benchmark (OGB) datasets, including four molecular graph datasets and one protein-protein association network.

Method	ogbg-molhiv	ogbg-moltox21	ogbg-moltoxcast	ogbg-ppa	ogbg-molpcba
GIN	75.58±1.40	74.91±0.51	63.41±0.74	68.92±1.00	22.66±0.28
GIN+VN	75.20±1.30	76.21±0.82	66.18±0.68	70.37±1.07	27.03±0.23
GSN	77.99±1.00	-	-	-	-
PNA	79.05±1.30	-	-	-	28.38±0.35
ID-GNN	78.30±2.00	-	-	-	-
Deep LRP	77.19±1.40	-	-	-	-
GraphSNN	78.51±1.70	75.45±1.10	65.40±0.71	70.66±1.65	24.96±1.50
GraphSNN+VN	**79.72±1.83**	**76.78±1.27**	**67.68±0.92**	**72.02±1.48**	**28.50±1.68**

Table: Classification accuracy on large graph classification.

- Classification w.r.t GraphSNN$_M$ models by replacing GCN, GAT, GIN, and GraphSAGE aggregation schemes by our aggregation scheme.

Method	Cora	Citeseer	Pubmed	NELL	ogbn-arxiv
GCN	81.5 ± 0.4	70.3 ± 0.5	79.0 ± 0.5	66.0 ± 1.7	71.74 ± 0.29
GraphSNN$_{GCN}$	**83.1 ± 1.8**	**72.3 ± 1.5**	**79.8 ± 1.2**	**68.3 ± 1.6**	**72.20 ± 0.90**
GAT	83.0 ± 0.6	72.6 ± 0.6	78.5 ± 0.3	-	-
GraphSNN$_{GAT}$	**83.8 ± 1.2**	**73.5 ± 1.6**	**79.6 ± 1.4**	-	-
GIN	77.6 ± 1.1	66.1 ± 1.5	77.0 ± 1.2	61.5 ± 2.3	-
GraphSNN$_{GIN}$	**79.2 ± 1.7**	**68.3 ± 1.5**	**78.8 ± 1.3**	**63.8 ± 2.7**	-
GraphSAGE	79.2 ± 3.7	71.6 ± 1.9	77.4 ± 2.2	63.7 ± 5.2	71.49 ± 0.27
GraphSNN$_{GraphSAGE}$	**80.5 ± 2.5**	**72.7 ± 3.2**	**79.0 ± 3.5**	**66.3 ± 5.6**	**71.80 ± 0.70**

Table: Classification accuracy on semi-supervised node classification.

- Oversmoothing analysis of GCN and GraphSNN$_{GCN}$ on the datasets Cora, Citeseer and Pubmed.