How Powerful are K-hop Message Passing Graph Neural Networks

Jiarui Feng, Yixin Chen, Fuhai Li, Anindya Sarkar, Muhan Zhang

February 5, 2025

Washington University; Peking University; Beijing Institute for General Artificial Intelligence

Presented by: George Giapitzakis

Setup

- Graph $G = (V, E)$. Vertex set $V = \{1, 2, \ldots, n\}$, edge set $E \subseteq V \times V$ and adjacency matrix $A \in \{0, 1\}^{n \times n}$.
- x_u: feature of node u.
- e_{uv}: feature of edge joining u and v.
- $Q^1_{v,G}$: set of 1-hop neighbors of node v in G and $\mathcal{N}^1_{v,G} \triangleq \{v\} \cup Q^1_{v,G}$.
- 1-**hop message passing framework**:

$$m_v^\ell = \mathrm{MES}^\ell\left(\left\{\!\!\left\{(h_u^{l-1}, e_{uv}) \mid u \in Q^1_{v,G}\right\}\!\!\right\}\right), h_v^\ell = \mathrm{UPD}^\ell(m_v^\ell, h_v^{\ell-1})$$

and

$$h_G = \mathrm{READOUT}\left(\left\{\!\!\left\{h_v^L \mid v \in V\right\}\!\!\right\}\right)$$

Setup

- ➔ **Shortest path distance (spd) kernel**
  - ⋆ $\mathcal{N}^{k,spd}_{v,G}$: set of nodes at distance **at most** k from v.
  - ⋆ $Q^{k,spd}_{v,G}$: set of nodes at distance **exactly** k from v.
- ➔ **Graph diffusion (gd) kernel**
  - ⋆ $\mathcal{N}^{k,gd}_{v,G}$: set of nodes that can diffuse information to v **within** k walk steps from v.
  - ⋆ $Q^{k,spd}_{v,G}$: set of nodes that can diffuse information to v **with** k walk steps from v.
- ➔ K**-hop message passing framework** $(t \in \{spd, gd\})$:

$$m_v^{\ell,k} = \mathrm{MES}_k^{\ell}\left(\left\{\!\!\left\{(h_u^{\ell-1}, e_{uv}) \mid u \in Q_{v,G}^{k,t}\right\}\!\!\right\}\right)$$

$$h_v^{\ell,k} = \mathrm{UPD}_k^{\ell}(m_v^{\ell,k}, h_v^{\ell-1})$$

$$h_v^{\ell} = \mathrm{COMBINE}^{l}\left(\left\{\!\!\left\{h_v^{\ell,k} \mid k = 1, 2, \ldots, K\right\}\!\!\right\}\right)$$

Example

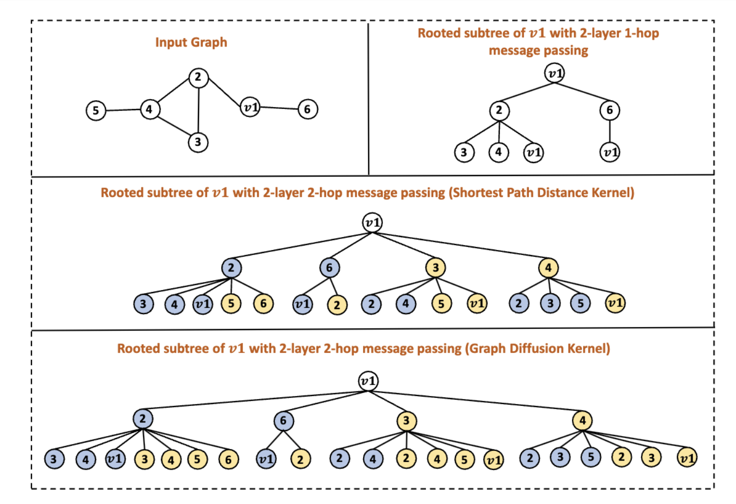

Figure 1: Example of 1-hop and K-hop neighbors for *spd* and *gd* kernels ($K = 2$).

Theoretical results

- **Definition.** A *proper* K-hop message passing GNN is a GNN model where the message, update, and combine functions are all injective given the input from a countable space.
- **Proposition 1.** A proper K-hop message passing GNN is strictly more powerful than 1-hop message passing GNNs when $K > 1$.
- **Theorem 1.** Consider all pairs of n-sized r-regular graphs, let $3 \leq r < (2 \log 2n)^{1/2}$ and ϵ be a fixed constant. With at most

$$K = \left\lfloor \left(\frac{1}{2} + \epsilon \right) \frac{\log 2n}{\log(r-1)} \right\rfloor,$$

there exists a 1 layer K-hop message passing GNN using the shortest path distance kernel that distinguishes almost all $1 - o(n^{-1/2})$ such pairs of graphs.

➔ **Theorem 2.** The expressive power of a proper K-hop message passing GNN of any kernel is bounded by the 3-WL test.

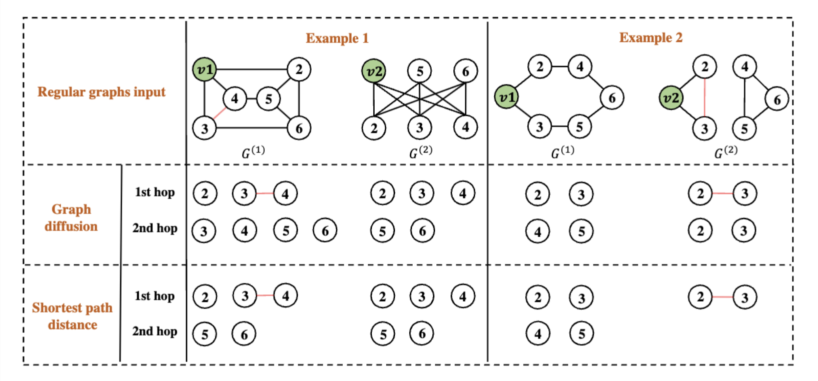

Figure 2: Example demonstrating Theorem 1.

KP-GNN: Improving K-hop message passing

- $E(Q_{v,G}^{k,t})$: set of edges connecting nodes from $Q_{v,G}^{k,t}$.
- $G_{v,G}^{k,t} = (Q_{v,G}^{k,t}, E(Q_{v,G}^{k,t}))$: *peripheral subgraph* induced by $Q_{v,G}^{k,t}$.
- **KP-GNN:** Augment MES with peripheral subgraph information:

$$m_v^{\ell,k} = \mathrm{MES}_k^{\ell}\left(\left\{\!\!\left\{(h_u^{\ell-1}, e_{uv}) \mid u \in Q_{v,G}^{k,t}\right\}\!\!\right\}, G_{v,G}^{k,t}\right)$$

- In particular,

$$\mathrm{MES}_k^{\ell} = \mathrm{MES}_k^{\ell,\text{normal}}\left(\left\{\!\!\left\{(h_u^{\ell-1}, e_{uv}) \mid u \in Q_{v,G}^{k,t}\right\}\!\!\right\}\right) + f(G_{v,G}^{k,t})$$

where

$$f(G_{v,G}^{k,t}) = \mathrm{EMB}((E(Q_{v,G}^{k,t}), C_k^{k'}))$$

Analysis of KP-GNN

- **Proposition 2** (omitted) shows that there exist pairs of isomorphic graphs that 3-WL fails to distinguish but KP-GNN can do so.
- Space complexity: $\mathcal{O}(n)$ for both K-hop GNN and KP-GNN with *spd* kernel.
- Time complexity: $\mathcal{O}(n^2)$ for both K-hop GNN and KP-GNN with *spd* kernel (compare to $\mathcal{O}(m)$ for MPNN).
- **Limitation:** Increased receptive field (use information from K-hop neighbors instead of 1-hop) hurts learning.

Experiments

- Datasets:
  - Synthetic: EXP, SR25, CSL.
  - Real-world: MUTAG, QM9, ZINC.
- Results:
  - K-hop GNNs outperform 1-hop GNNs on synthetic datasets.
  - KP-GNN achieves superior results on distinguishing non-isomorphic graphs and graph property prediction.
  - Competitive performance on real-world datasets.

Experiments with synthetic data

Method	K	EXP (ACC)		SR (ACC)		CSL (ACC)	
		SPD	GD	SPD	GD	SPD	GD
K-GIN	K=1	50	50	6.67	6.67	12	12
	K=2	50	50	6.67	6.67	32	22.7
	K=3	100	66.9	6.67	6.67	62	42
	K=4	100	100	6.67	6.67	92.7	62.7
KP-GIN	K=1	50	50	100	100	22	22
	K=2	100	100	100	100	52.7	52.7
	K=3	100	100	100	100	90	90
	K=4	100	100	100	100	100	100

(a) Expressivity experiment.

Method	Node Properties ($\log_{10}$(MSE))			Graph Properties ($\log_{10}$(MSE))			Counting Substructures (MAE)			
	SSSP	Ecc.	Lap.	Connect.	Diameter	Radius	Tri.	Tailed Tri.	Star	4-Cycle
GIN	-2.0000	-1.9000	-1.6000	-1.9239	-3.3079	-4.7584	0.3569	0.2373	0.0224	0.2185
PNA	-2.8900	-2.8900	-3.7700	-1.9395	3.4382	-4.9470	0.3532	0.2648	0.1278	0.2430
PPGN	-	-	-	-1.9804	-3.6147	-5.0878	**0.0089**	**0.0096**	0.0148	0.0090
GIN-AK+	-	-	-	**-2.7513**	**-3.9687**	-5.1846	0.0123	0.0112	**0.0150**	**0.0126**
K-GIN+	-2.7919	-2.5938	**-4.6360**	-2.1782	-3.9695	**-5.3088**	0.2593	0.1930	0.0165	0.2079
KP-GIN+	**-2.7969**	**-2.6169**	-4.7687	-4.4322	-3.9361	-5.3345	0.0060	0.0073	0.0151	0.0395

(b) Graph properties experiment.

Experiments with real-world data

Method	MUTAG	D&D	PTC-MR	PROTEINS	IMDB-B
WL	90.4±5.7	79.4±0.3	59.9±4.3	75.0±3.1	73.8±3.9
GIN	89.4±5.6	-	64.6±7.0	75.9±2.8	75.1±5.1
DGCNN	85.8±1.7	79.3 ±0.9	58.6 ±2.5	75.5±0.9	70.0±0.9
GraphSNN	91.24±2.5	82.46±2.7	66.96±3.5	76.51±2.5	76.93±3.3
GIN-AK+	91.30±7.0	-	68.20±5.6	77.10±5.7	75.60±3.7
KP-GCN	91.7±6.0	79.0±4.7	67.1±6.3	75.8±3.5	75.9±3.8
KP-GraphSAGE	91.7±6.5	78.1±2.6	66.5±4.0	76.5±4.6	76.4±2.7
KP-GIN	92.2±6.5	79.4±3.8	66.8±6.8	75.8±4.6	76.6±4.2
GIN-AK+*	95.0±6.1	OOM	74.1±5.9	78.9±5.4	77.3±3.1
GraphSNN*	94.70±1.9	**83.93±2.3**	70.58±3.1	78.42±2.7	78.51±2.8
KP-GCN*	**96.1±4.6**	83.2±2.2	**77.1±4.1**	80.3±4.2	79.6±2.5
KP-GraphSAGE*	**96.1±4.6**	83.6±2.4	76.2±4.5	**80.4±4.3**	80.3±2.4
KP-GIN*	95.6±4.4	83.5±2.2	76.2±4.5	79.5±4.4	**80.7±2.6**

(a) TU dataset.

Method	# param.	test MAE
MPNN	480805	0.145±0.007
PNA	387155	0.142±0.010
Graphormer	489321	0.122±0.006
GSN	~500000	0.101±0.010
GIN-AK+	-	0.080±0.001
CIN	-	**0.079±0.006**
KP-GIN+	499099	0.111±0.006
KP-GIN′	488649	0.093±0.007

(b) ZINC dataset.

Target	**DTNN**	**MPNN**	**Deep LRP**	**PPGN**	**N-1-2-3-GNN**	**KP-GIN+**	**KP-GIN′**
μ	**0.244**	0.358	0.364	**0.231**	0.433	0.367	0.358
α	0.95	0.89	0.298	0.382	0.265	**0.242**	**0.233**
$\varepsilon_{\text{HOMO}}$	0.00388	0.00541	0.00254	0.00276	0.00279	**0.00247**	**0.00240**
$\varepsilon_{\text{LUMO}}$	0.00512	0.00623	0.00277	0.00287	0.00276	**0.00238**	**0.00236**
$\Delta\varepsilon$	0.0112	0.0066	0.00353	0.00406	0.00390	**0.00345**	**0.00333**
$\langle R^2 \rangle$	17.0	28.5	19.3	16.7	20.1	**16.49**	**16.51**
ZPVE	0.00172	0.00216	0.00055	0.00064	**0.00015**	0.00018	**0.00017**
U_0	2.43	2.05	0.413	0.234	0.205	**0.0728**	**0.0682**
U	2.43	2.00	0.413	0.234	0.200	**0.0553**	**0.0696**
H	2.43	2.02	0.413	0.229	0.249	**0.0575**	**0.0641**
G	2.43	2.02	0.413	0.238	0.253	**0.0526**	**0.0484**
C_v	0.27	0.42	0.129	0.184	**0.0811**	0.0973	**0.0869**

(c) QM9 dataset.

Questions?

Thank you!
Any Questions?