HOW POWERFUL ARE K-HOP MESSAGE PASSING GRAPH NEURAL NETWORKS

Present: Qianqiu Zhang

Author: Feng et al.

OUTLINE

- Message Passing
  - 1-hop
  - K-hop
    - shortest path distance (spd) kernel
    - graph diffusion (gd) kernel
- Expressive power of K-hop
- Limitation of K-hop
- KP-GNN
- Experiment
- Issues

GRAPH NOTATIONS

- Q: denotes the set of nodes in G that are exactly the k-th hop neighbors of v
  - 1-hop neighbors: $\mathcal{N}^1_{v,G} = Q^1_{v,G} \cup \{v\}$
  - k-hop neighbors: ALL the neighbors that have distance from node v less than or equal to K
  - k-th hop neighbors: neighbors with EXACTLY distance k from node v
- Regular graph: In graph theory, a regular graph is a graph where each vertex has the same number of neighbors; i.e. every vertex has the same degree.

OVERVIEW

- The paper raise the issue that existing literatures are upper bound by 3-WL test
  - 1-hop GNN is upper bound by 1-WL test
  - K-hop upper bound by 3-WL test
- They propose a new feature added to the k-hop GNN to increase the power of the model

WHAT IS THE PROBLEM

- 1-hop Message Passing

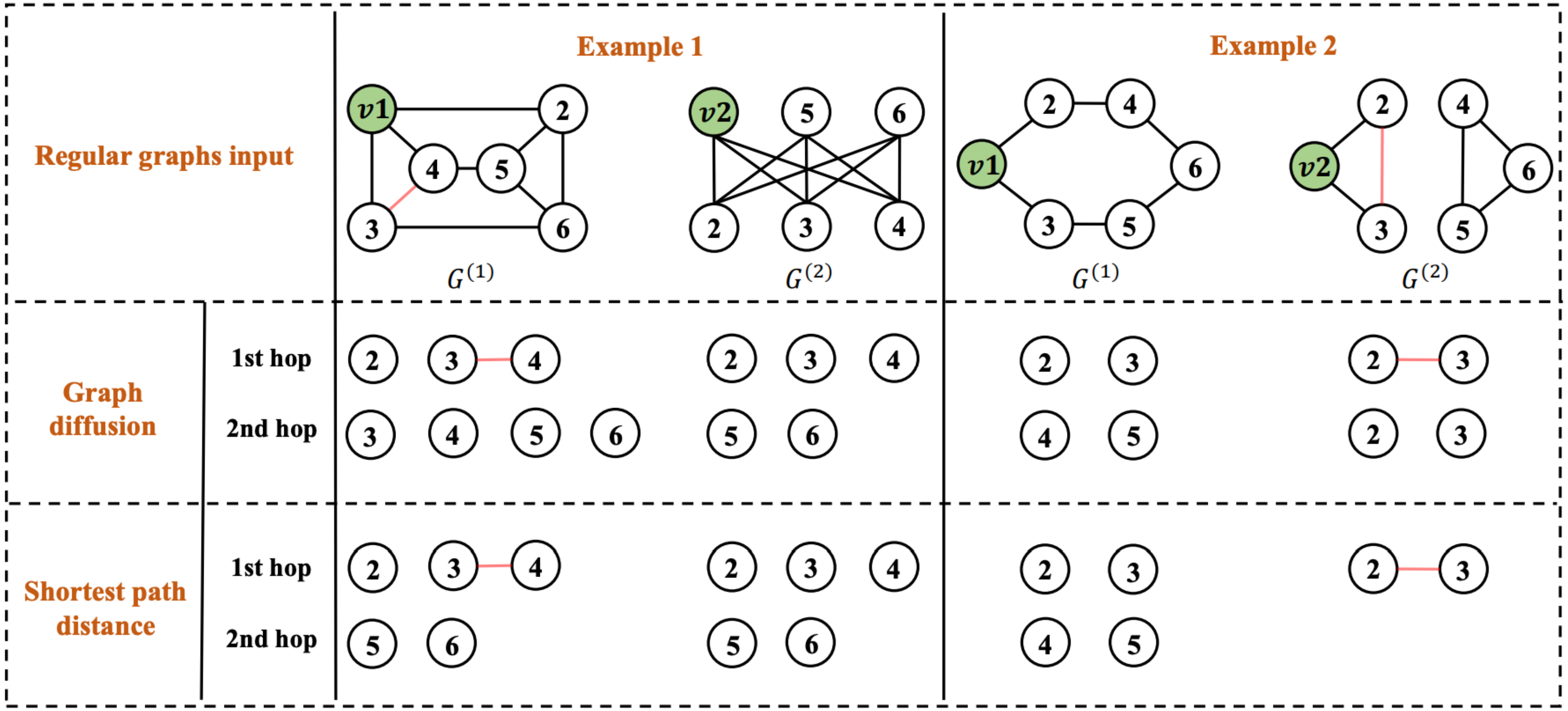
$$m_v^l = \mathbf{MES}^l(\{\!\{(h_u^{l-1}, e_{uv}) | u \in Q_{v,G}^1\}\!\}), \quad h_v^l = \mathbf{UPD}^l(m_v^l, h_v^{l-1}),$$

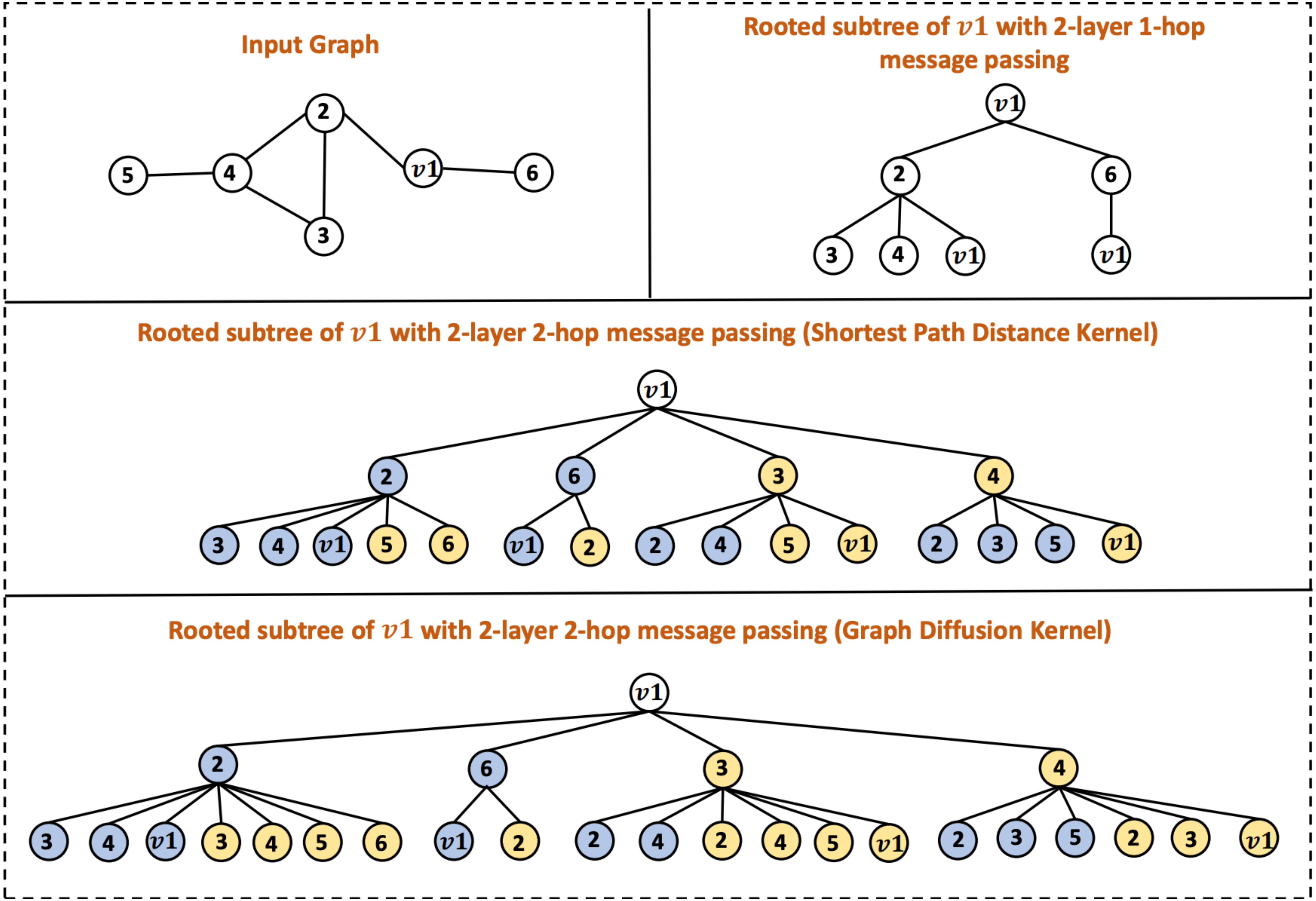
- Expressive power of 1-hop GNN
  - 1-hop GNN Upper bounded by 1-WL
    - Mimic higher-order WL-test increase time complexity exponentially

NEIGHBORHOOD DEFINITION

- K-hop
  - shortest path distance (spd) kernel
    - The set of nodes from node v that have the shortest path distance less than or equal to K
  - graph diffusion (gd) kernel
    - The set of nodes that can diffuse information to node v within the number of random walk diffusion steps K

NEIGHBORHOOD DEFINITION




Input Graph
2
5
4
v1
6
3
Rooted subtree of v1 with 2-layer 1-hop message passing
v1
2
6
3
4
v1
v1
Rooted subtree of v1 with 2-layer 2-hop message passing (Shortest Path Distance Kernel)
v1
2
6
3
4
3
4
v1
5
6
v1
2
2
4
5
v1
2
3
5
v1
Rooted subtree of v1 with 2-layer 2-hop message passing (Graph Diffusion Kernel)
v1
2
6
3
4
3
4
v1
3
4
5
6
v1
2
2
4
2
4
5
v1
2
3
5
2
3
v1

RANDOM WALK VS SHORTEST PATH

- Random walk kernel can increase the number of neighbors at each iteration
- Node repetition can be harmful
- Higher time complexity

K-HOP

- K-hop Message Passing

$$m_v^{l,k} = \mathbf{MES}_k^l(\{\!\{(h_u^{l-1}, e_{uv}) | u \in Q_{v,G}^{k,t})\}\!\}), \quad h_v^{l,k} = \mathbf{UPD}_k^l(m_v^{l,k}, h_v^{l-1}),$$
$$h_v^l = \mathbf{COMBINE}^l(\{\!\{h_v^{l,k} | k = 1, 2, ..., K\}\!\}),$$

$Q_{v,G}^{k,t}$ denotes neighbors of node v in graph G within K hops under t kernel

- K-hop Message Passing GNN
  - Message, update, and combine functions are all injective

FEATURE CONSTRUCTION

- Assume all node features only depends on the graph structure
- node configuration:

$$A_{v,G}^{K,t} = (a_{v,G}^{1,t}, a_{v,G}^{2,t}, \ldots, a_{v,G}^{K,t}), \textit{ where } a_{v,G}^{i,t} = |Q_{v,G}^{i,t}|$$

EXISTING STATE-OF-THE-ART

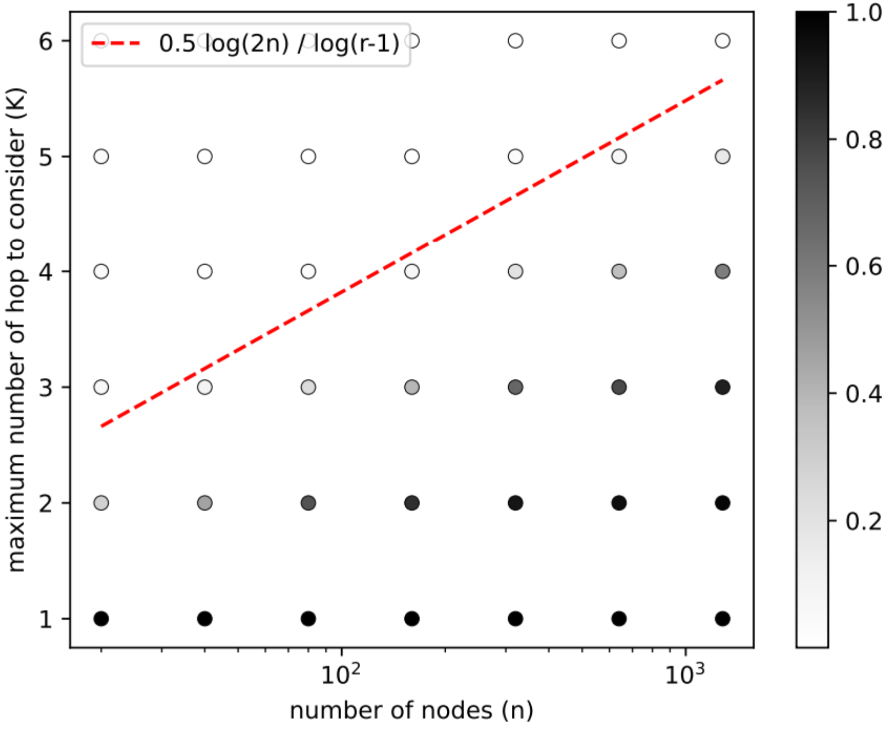
- K-hop message passing GNN is strictly more powerful than 1-hop message passing GNNs when K > 1.

Theorem 1. *Consider all pairs of n-sized r-regular graphs, let $3 \leq r < (2log2n)^{1/2}$ and ϵ be a fixed constant. With at most $K = \lfloor (\frac{1}{2} + \epsilon) \frac{\log 2n}{\log (r-1)} \rfloor$, 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.*

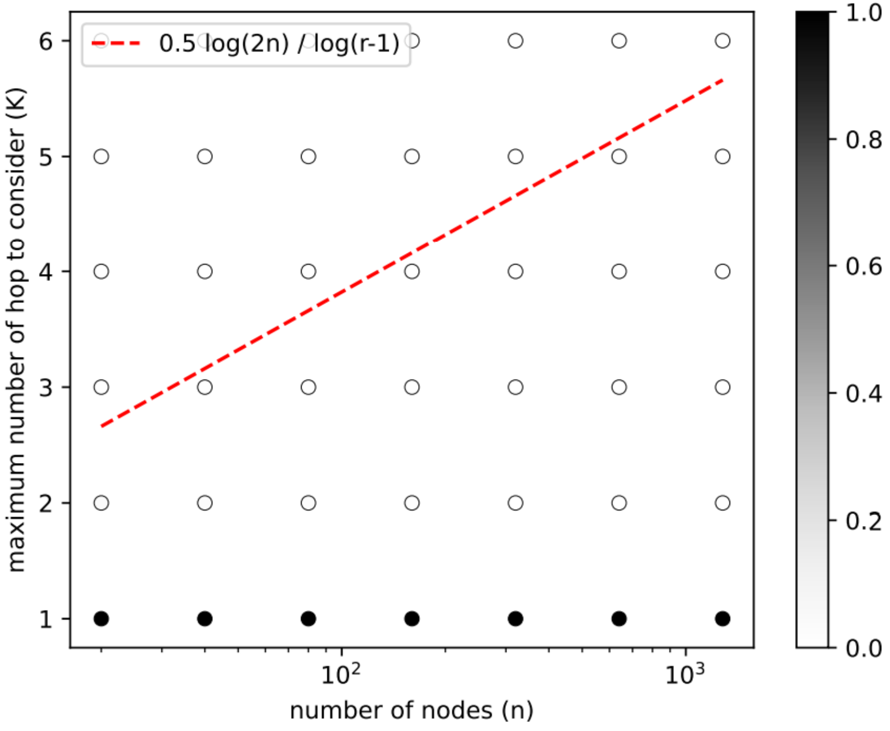
EXISTING STATE-OF-THE-ART

- N-sized r-regular graph $3 \leq r < (2log2n)^{1/2}$
- K-hop distance $K = \lfloor (\frac{1}{2} + \epsilon) \frac{\log 2n}{\log (r-1)} \rfloor$
- High chance $1 - o(n^{-1/2})$ GNN can distinguish non-isomorphic graphs

EXISTING STATE-OF-THE-ART



Node level



Graph level

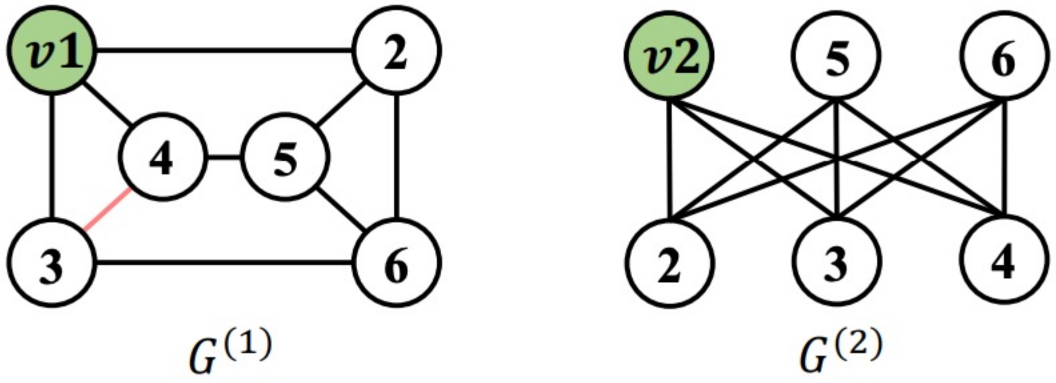
WHY DON'T PREVIOUS METHODS WORK ON THAT PROBLEM?

- K-hop message passing GNN limitation
  - the expressive power of K-hop message passing is bounded by 3-WL.
  - cannot distinguish any non-isomorphic distance regular graphs

WHAT IS THE SOLUTION TO THE PROBLEM THE AUTHORS PROPOSE?

Add peripheral edge into message passing can help distinguish between regular graphs even in 1-hop setting !

- K
  - peripheral edge: edges connect nodes within set $Q^{k,t}_{v,G}$
  - peripheral subgraph is a graph constituted by peripheral edges



$$E(Q^{1,t}_{v_1,G^{(1)}}) = \{(3,4)\} \quad E(Q^{1,t}_{v_2,G^{(2)}}) = \{\}$$

KP-GNN

- K-hop Message Passing

$$m_v^{l,k} = \text{MES}_k^l(\{\!\{(h_u^{l-1}, e_{uv})|u \in Q_{v,G}^{k,t})\}\!\}), \quad h_v^{l,k} = \text{UPD}_k^l(m_v^{l,k}, h_v^{l-1}),$$
$$h_v^l = \text{COMBINE}^l(\{\!\{h_v^{l,k}|k = 1, 2, ..., K\}\!\}),$$

- KP-GNN Message passing

$$m_v^{l,k} = \text{MES}_k^l(\{\!\{(h_u^{l-1},\ e_{uv})|u \in Q_{v,G}^{k,t}\}\!\},\ G_{v,G}^{k,t}).$$
$$\text{MES}_k^l = \text{MES}_k^{l,normal}(\{\!\{(h_u^{l-1}, e_{uv})|u \in Q_{v,G}^{k,t}\}\!\}) + f(G_{v,G}^{k,t}),$$
$$f(G_{v,G}^{k,t}) = \text{EMB}((E(Q_{v,G}^{k,t}), C_k^{k'}))\ ,$$

KP-GNN

$$
\begin{aligned}
\mathrm{MES}_k^l &= \mathrm{MES}_k^{l,normal}(\{\!\{(h_u^{l-1}, e_{uv}) | u \in Q_{v,G}^{k,t}\}\!\}) + f(G_{v,G}^{k,t}), \\
f(G_{v,G}^{k,t}) &= \mathrm{EMB}((E(Q_{v,G}^{k,t}), \boxed{C_k^{k'}})) \ ,
\end{aligned}
$$

$C_k^{k'}$ Encode both node configuration and number of peripheral edges of all nodes in the peripheral subgraph

KP-GNN

After obtaining message

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

Combining node representation from different hops

$$h_v^l = \text{COMBINE}^l(\{\!\{h_v^{l,k} | k = 1, 2, ..., K\}\!\})$$

Compute graph representation

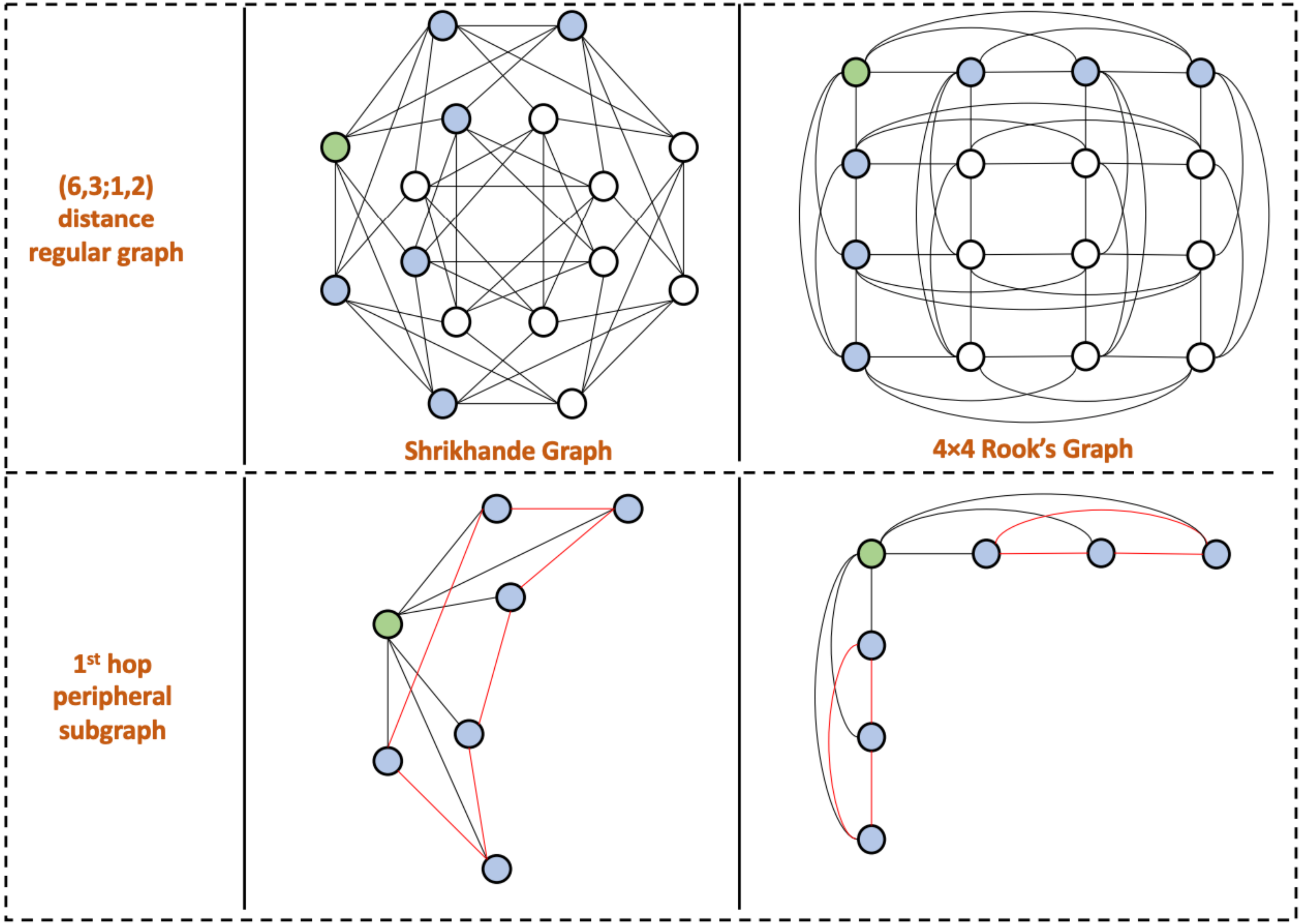
$$h_G = \text{READOUT}(\{\!\{h_v^L | v \in V\}\!\})$$

KP-GNN

- K-hop GNNs cannot distinguish any non-isomorphic distance regular graphs
- Analyze KP-GNN's ability of distinguishing distance regular graphs

WH

- The



(6,3;1,2)
distance
regular graph
Shrikhande Graph
4×4 Rook's Graph
1st hop
peripheral
subgraph

WHY KP-GNN

Proposition 2. *For two non-isomorphic distance regular graphs* $G^{(1)} = (V^{(1)}, E^{(1)})$ *and* $G^{(2)} = (V^{(2)}, E^{(2)})$ *with the same diameter* d *and intersection array* $(b_0, b_1, ..., b_{d-1}; c_1, c_2, ..., c_d)$*. Given a proper 1-layer* d*-hop KP-GNN with message functions defined in Equation (5), it can distinguish* $G^{(1)}$ *and* $G^{(2)}$ *if* $C^{k'}_{j,G^{(1)}} \neq C^{k'}_{j,G^{(2)}}$ *for some* $0 < j \leq d$*.*

TIME COMPLEXITY

- K-hop message passing GNNs: $O(n^2)$
- KP-GNN: extra time complexity comes from counting the peripheral edges and k'–configuration *(can be preprocessed)*

$$O(n^2)$$

EXPERIMENT

- Datasets
  - Determine isomorphism
    - EXP datasets: contains 600 pairs of non-isomorphic graphs (1-WL failed).
    - SR25 dataset: 15 non-isomorphic strongly regular graphs (3- WL failed)
    - CSL dataset, which contains 150 4-regular graphs (1-WL failed) divided into 10 isomorphism classes

DETERMINE ISOMORPHISM

- Use encoder of GIN

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

GRAPH/NODE PROPERTY REGRESSION

- Regression test

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

REAL-WORLD EXPERIMENT

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**

REAL-WORLD EXPERIMENT

- Molecule prediction test

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

ISSUE

- The experiment on synthetic graph prove the theorem
- It does not consistently improve in real world applications
- Using k-hop will increase the receptive field of a node w.r.t K
  - After L layer, the receptive field of a node is $K \times L$
    - KP-GNN' at the first layer but 1-hop message passing at the rest of the layers.

THANK YOU