Supplemental Material of MGNN

Xuexin Chen, Ruichu Cai*, Yuan Fang*, Min Wu*, Zijian Li, Zhifeng Hao

I. Proof for Lemma 1

Proof. First, we show the relationship between the graph's adjacency matrix and the motif-based adjacency matrix. Then, using this relationship, we finish the proof of the lemma.

On the directed graph G with self-loops, the subgraph composed of any node linked to any two of its neighbors is always an instance of open motif (M_8–M_{13}). That is, in the adjacency matrix $\mathbf{A}$ of the graph with self-loops, if $(\mathbf{A})_{ij} > 0$, $(\mathbf{A})_{uv} > 0$ where (i,j) and (u,v) are adjacent edges in G, then there always exist $k' \in \{8,9,...,13\}$ such that $\mathbf{A}_{k'}$ satisfies $(\mathbf{A}_{k'})_{ij} > 0$, $(\mathbf{A}_{k'})_{uv} > 0$. It immediately follows that, on a graph with self-loops, if $(\mathbf{A})_{ij} > 0$, then we also have $(\mathbf{A}_{k'})_{ij} > 0$. Without loss of generality, we assume $k' = 13$ for ease of discussion later. That is, $\forall (i,j) \in \mathcal{E}$, $(\mathbf{A}_{13})_{ij} > 0$.

Next, we use the construction method to complete the proof of this lemma. Based on Table I, an instance of standard GNN is in the form of $\tilde{\mathbf{h}}_v^{(l)} = \sigma(\omega\left(\left\{(\mathbf{A})_{vi}\mathbf{W}_s^{(l)}\tilde{\mathbf{h}}_i^{(l-1)} \middle| i \in \mathcal{N}(v)\right\}\right))$. We will use the following steps to find a special case of MGNN which have the same representational capacity as standard GNN.

First, this special case of MGNN must satisfy the following equation.

$$\begin{aligned}&\|_{k=1}^{13}\sigma(\omega(\{\alpha_{k,vi}^{(l)}\cdot(\mathbf{A}_k)_{vi}\mathbf{W}_m^{(l)}\mathbf{h}_i^{(l-1)}|i\in\mathcal{N}(v)\}))\\&=\sigma(\omega(\{\|_{k=1}^{12}\mathbf{0}_k\|(\mathbf{A})_{vi}\mathbf{W}_s^{(l)}\tilde{\mathbf{h}}_i^{(l-1)}|i\in\mathcal{N}(v)\})),\end{aligned}\tag{S1}$$

where $\mathbf{0}_k$ is a d_l-dimensional zero vector, $\mathbf{W}_m^{(l)}$ and $\mathbf{W}_s^{(l)} \in \mathbb{R}^{d_l \times d_{l-1}}$, $\mathbf{h}_i^{(l-1)}$ and $\tilde{\mathbf{h}}_i^{(l-1)} \in \mathbb{R}^{d_{l-1}}$, so that the dimensions on both sides of Eq. (S1) are the same. That is, the output dimensions of the special case of MGNN and standard GNN are the same, both being $13d_l$.

Next, with $\mathbf{W}_m^{(l)}$ and $\alpha_{k,vi}^{(l)}$ as variables, our goal is to prove that there will always be solutions to $\mathbf{W}_m^{(l)}$ and $\alpha_{k,vi}^{(l)}$ such that Eq. (S1) holds.

For simplicity, in Eq. (S1), we use symbol φ, a aggregation function with activation, to represent $\sigma \circ \omega$, that is,

$$\begin{aligned}&\|_{k=1}^{13}\varphi(\{\alpha_{k,vi}^{(l)}\cdot(\mathbf{A}_k)_{vi}\mathbf{W}_m^{(l)}\mathbf{h}_i^{(l-1)}|i\in\mathcal{N}(v)\})\\&=\varphi(\{\|_{k=1}^{12}\mathbf{0}_k\|(\mathbf{A})_{vi}\mathbf{W}_s^{(l)}\tilde{\mathbf{h}}_i^{(l-1)}|i\in\mathcal{N}(v)\}).\end{aligned}\tag{S2}$$

In the left hand side (LHS) of Eq. (S2), the result will not change if the order of concatenation operation and aggregation φ is exchanged. This is because the result value for each dimension in the LHS is only aggregated from the values of the same dimension in different feature vectors, and each feature vector is completely preserved after concatenation is performed. Thus, the LHS of Eq. (S2) becomes

$$\varphi(\{\|_{k=1}^{13}\alpha_{k,vi}^{(l)}\cdot(\mathbf{A}_k)_{vi}\mathbf{W}_m^{(l)}\mathbf{h}_i^{(l-1)}|i\in\mathcal{N}(v)\}).\tag{S3}$$

By combining Eq. (S2)–(S3), we get the equivalent form of Eq. (S1):

$$\begin{aligned}&\varphi(\{\|_{k=1}^{13}\alpha_{k,vi}^{(l)}\cdot(\mathbf{A}_k)_{vi}\mathbf{W}_m^{(l)}\mathbf{h}_i^{(l-1)}|i\in\mathcal{N}(v)\})\\&=\varphi(\{\|_{k=1}^{12}\mathbf{0}_k\|(\mathbf{A})_{vi}\mathbf{W}_s^{(l)}\tilde{\mathbf{h}}_i^{(l-1)}\}|i\in\mathcal{N}(v)\}).\end{aligned}\tag{S4}$$

Therefore, our goal now is to prove that there will always be solutions such that Eq. (S4) holds. We can solve for the following Eqs. (S5)–(S6) to ensure that Eq. (S4) holds. For $k \in \{1,...,12\}$,

$$\alpha_{k,vi}^{(l)}\cdot(\mathbf{A}_k)_{vi}\mathbf{W}_m^{(l)}\mathbf{h}_i^{(l-1)} = \mathbf{0}_k,\tag{S5}$$

and for $k = 13$,

$$\alpha_{13,vi}^{(l)}\cdot(\mathbf{A}_{13})_{vi}\mathbf{W}_m^{(l)}\mathbf{h}_i^{(l-1)} = (\mathbf{A})_{vi}\mathbf{W}_s^{(l)}\tilde{\mathbf{h}}_i^{(l-1)}.\tag{S6}$$

Then we will demonstrate that $\forall l \geq 1$, there will always be solutions to $\mathbf{W}_m^{(l)}$ and $\alpha_{k,vi}^{(l)}$, such that Eqs. (S5)–(S6) holds. Specifically, when $l = 1$, $\mathbf{h}_i^{(0)} = \tilde{\mathbf{h}}_i^{(0)} = \mathbf{x}_i$, allowing Eqs. (S5)–(S6) to hold for $\mathbf{W}_m^{(l)} = \frac{(\mathbf{A})_{vi}}{\alpha_{13,vi}^{(l)}\cdot(\mathbf{A}_{13})_{vi}}\mathbf{W}_s^{(l)}$, $\alpha_{13,vi}^{(l)} \neq 0$ and $\alpha_{k,vi}^{(l)} = 0$ ($k \in \{1,...,12\}$), that is, 1-th special case of MGNN layer can generate the same vector representation as 1-th standard GNN layer since both models have the same output in the previous layer (i.e., $\mathbf{h}_i^{(0)} = \tilde{\mathbf{h}}_i^{(0)}$). Similarly, when $l > 1$, Eqs. (S5)–(S6) holds. This finishes the proof of the lemma. □

II. Performance Evaluation of Ensemble GNNs

We evaluate the empirical performance of MGNN against ensemble GNNs and standard GNNs in Table S1 and Table S2.

As shown in Table S1, MGNN significantly and consistently outperforms all the baselines on different datasets. In particular, ESAGE achieves the second best performance on Pubmed, while EGAT+SAGE achieves the second best performance on Cora and Chem2Bio2RDF. On Citeseer, GAT achieves the second best performance. MGNN is able to achieve further improvements against ESAGE by 2.12% on Pubmed, against GAT by 1.22% on Citeseer, as well as against EGAT+SAGE by 3.05% and 2.14% on Cora and Chem2Bio2RDF respectively.

In Table S2, similarly, MGNN regularly surpasses all baselines. In particular, ESAGE achieves the second best performance on ENZYMES, while GraphSAGE achieves the second best performance on MUTAG and GCN achieves the second best performance on AIDS. Our MGNN is capable of achieving further improvements against ESAGE by 14.98% on ENZYMES, as well as against GraphSAGE and GCN by 5.96% on MUTAG and by 0.76% on AIDS, respectively.

TABLE S1
PERFORMANCE COMPARISON ON THE NODE CLASSIFICATION TASK, MEASURED IN ACCURACY. STANDARD DEVIATION ERRORS ARE GIVEN.

	Cora	Citeseer	Pubmed	Chem2Bio2RDF
GCN	0.8595 ± 0.0207	0.7764 ± 0.0045	0.8865 ± 0.0048	0.9371 ± 0.0017
GraphSAGE	0.8610 ± 0.0101	0.7744 ± 0.0061	0.8980 ± 0.0049	0.9630 ± 0.0010
GAT	0.8775 ± 0.0127	0.7852 ± 0.0052	0.8840 ± 0.0079	0.9628 ± 0.0017
GIN	0.8107 ± 0.0188	0.7255 ± 0.0160	0.8810 ± 0.0156	0.9205 ± 0.0129
BGNN	0.8470 ± 0.0143	0.7750 ± 0.0112	0.8380 ± 0.0119	0.8746 ± 0.0115
EGAT	0.8720 ± 0.0040	0.7220 ± 0.0060	0.8970 ± 0.0010	0.9658 ± 0.0040
ESAGE	0.8612 ± 0.0135	0.7604 ± 0.0171	0.9040 ± 0.0102	0.9633 ± 0.0019
EGAT+SAGE	0.8792 ± 0.0102	0.7632 ± 0.0141	0.8992 ± 0.0109	0.9663 ± 0.0010
MGNN	**0.9060** ± 0.0049	**0.7948** ± 0.0050	**0.9232** ± 0.0084	**0.9870** ± 0.0021

TABLE S2
PERFORMANCE COMPARISON ON THE GRAPH CLASSIFICATION TASK, MEASURED IN ACCURACY. STANDARD DEVIATION ERRORS ARE GIVEN.

	MUTAG	ENZYMES	AIDS
GCN	0.7555 ± 0.0651	0.2100 ± 0.0285	0.9895 ± 0.0091
GAT	0.7391 ± 0.0315	0.1667 ± 0.0000	0.8740 ± 0.1013
GraphSAGE	0.7984 ± 0.0526	0.2333 ± 0.0586	0.9855 ± 0.0091
GIN	0.7780 ± 0.0940	0.2630 ± 0.0330	0.9870 ± 0.0090
EGAT	0.7820 ± 0.0610	0.2420 ± 0.0450	0.9850 ± 0.0050
ESAGE	0.7350 ± 0.0650	0.2670 ± 0.0560	0.9850 ± 0.0060
EGAT+SAGE	0.7340 ± 0.0320	0.2500 ± 0.0480	0.9840 ± 0.0070
MGNN	**0.8460** ± 0.0230	**0.3070** ± 0.0300	**0.9970** ± 0.0030

TABLE S3
THE EFFICIENCY ANALYSIS OF THREE METHODS FOR CONSTRUCTING MOTIF-BASED ADJACENCY MATRIX, IN TERMS OF THE RUNNING TIME (SECONDS). 'MATMUL' DENOTES MATRIX MULTIPLICATION METHOD.

	# Nodes	Closed Motif: M1	Open Motif: M13	
		MatMul [10]	Enumerate	Non-enumerate
Cora	2,708	0.003	73.322	1.534
Pubmed	19,717	0.027	4249.435	18.852
Chem2-Bio2RDF	295,911	0.228	1226K	69.353

III. EFFICIENCY ANALYSIS OF MOTIF-BASED ADJACENCY MATRIX CONSTRUCTION

We evaluate the efficiency of MatMul [1] for closed motifs and our proposed non-enumeration method for open motifs, in terms of the running time, in Table S3 below. For open motifs, we would compare the running time of both enumeration and non-enumeration methods.

As shown in Table S3, it can be observed that MatMul can run very fast for closed motifs even for large-scale graphs, such as Chem2Bio2RDF. Meanwhile, compared to the standard enumeration method, our proposed non-enumeration method performs much better for open motifs. Even for Chem2Bio2RDF dataset, our non-enumeration can still run quite fast, taking about 69 seconds to construct the adjacency matrix for the open motif M_{13}. These results demonstrate that our preprocessing for both closed and open motifs is efficient.

TABLE S4
PERFORMANCE AND EFFICIENCY ANALYSIS OF MGNN USING ALL MOTIFS OR NOT, MEASURED IN ACCURACY AND OVERALL TRAINING TIME (MINUTES). '(M7, M8, M9)' DENOTES THAT MGNN USES ONLY M_7, M_8 AND M_9 MOTIFS.

	# Nodes	ACC		Overall/min	
		(M7, M8, M9)	ALL	(M7, M8, M9)	ALL
Cora	2,708	0.8732	0.9060	0.87	1.37
CiteSeer	3,327	0.7224	0.7948	0.80	1.29
PubMed	19,717	0.4220	0.9232	5.73	10.00
Chem2-Bio2RDF	295,911	0.9741	0.9870	14.26	27.26

IV. PERFORMANCE AND EFFICIENCY ANALYSIS OF MGNN USING ALL MOTIFS

We compare the performance and efficiency of MGNN using all motifs or not, in terms of accuracy and overall training time. Specifically, we select motifs M_7, M_8 and M_9 which are commonly important in Cora, CiteSeer, PubMed and Chem2Bio2RDF, and make MGNN utilize just the above three motifs to conduct node classification on the four datasets. For simplicity, we denote this variant of MGNN as (M7, M8, M9).

As shown in Table S4, MGNN using all the motifs achieves better accuracy, while (M7, M8, M9) method can clearly save the training time. However, (M7, M8, M9) method achieves lower accuracy than MGNN using all the motifs on all four datasets, showing that these three motifs are not sufficient to capture all the important high-order structures for these four datasets. In addition, we would think the efficiency when using all the motifs is still satisfactory. Even on the largest dataset (i.e., Chem2Bio2RDF), the overall training time for MGNN using all the motifs is just 13 minutes longer than (M7, M8, M9) method, while on other datasets the differences are much smaller.

REFERENCES

[1] H. Zhao, X. Xu, Y. Song, D. L. Lee, Z. Chen, and H. Gao, "Ranking users in social networks with higher-order structures," in *Proc. Association for the Advancement of Artif. Intell.*, 2018, pp. 232–240.