Can Classic GNNs Be Strong Baselines for Graph-level Tasks? Simple Architectures Meet Excellence

Yuankai Luo, Lei Shi, Xiao-Ming Wu

ICML
International Conference
On Machine Learning

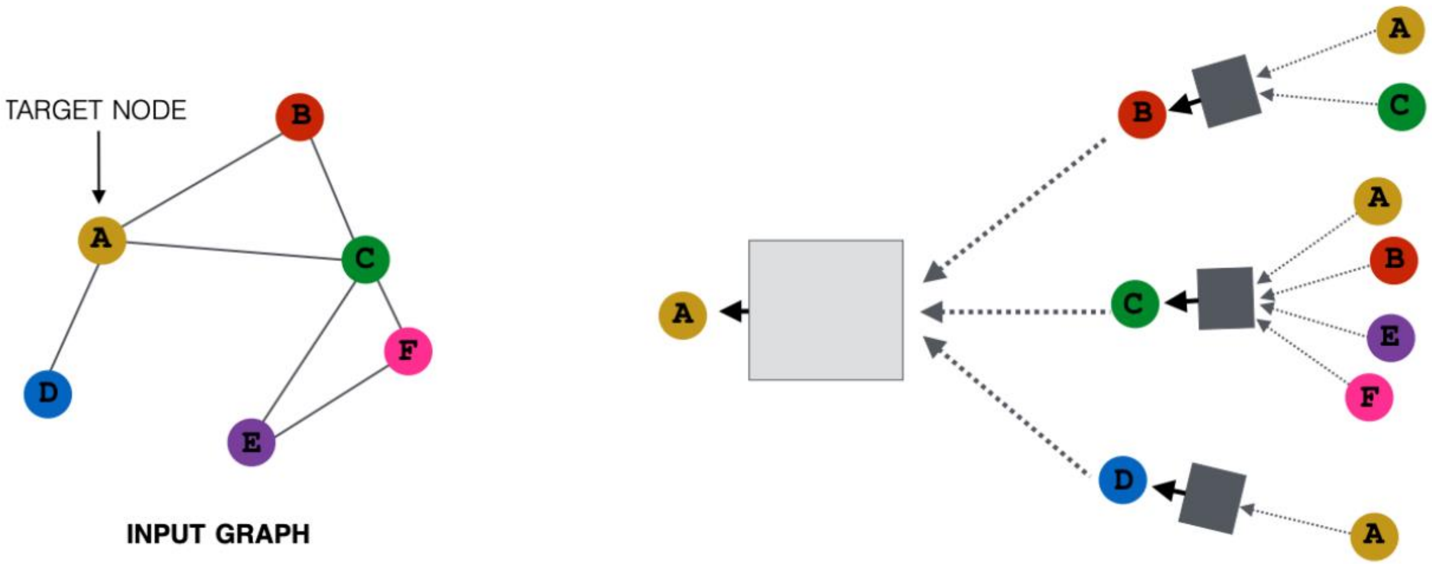
北京航空航天大學
BEIHANG UNIVERSITY

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Background

■ Message-Passing GNNs :

A message-passing mechanism creates new node representations, where each node gathers information from its neighbors and combines it to update its own embedding.

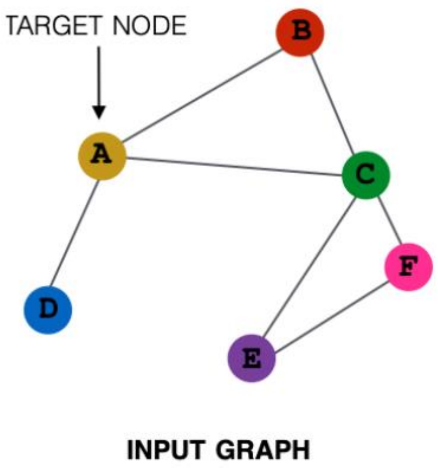


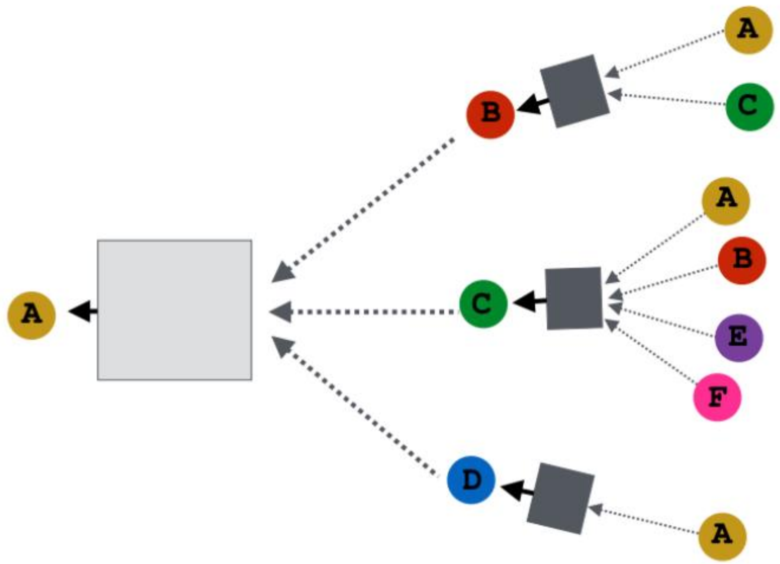
Background

Message-Passing GNNs :

A message-passing mechanism creates new node representations, where each node gathers information from its neighbors and combines it to update its own embedding.

Over-smoothing
Over-squashing
…





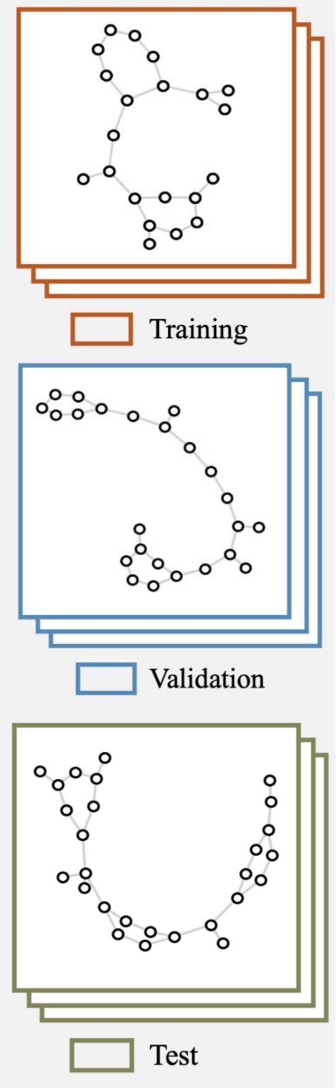
Graph-level Tasks

A graph-level dataset: $\Gamma = \{(G_i, y_i)\}_i$, each graph G_i is associated with a label vector y_i, representing either categorical labels for classification or continuous values for regression.

Next, the dataset Γ is typically split into training, validation, and test sets, denoted as $\Gamma = \Gamma_{train} \cup \Gamma_{val} \cup \Gamma_{test}$.

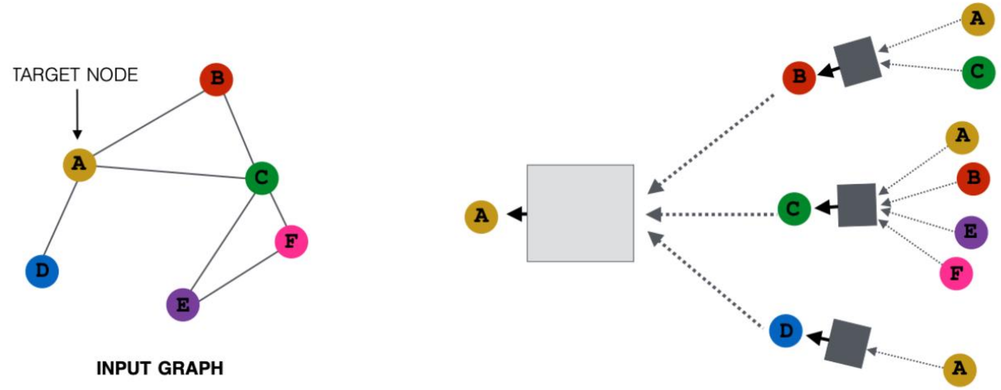
For each graph G_i, we apply a readout function $R(\cdot)$ (such as mean pooling) to aggregate the node representations produced by the final GNN layer, yielding a graph-level representation h_i^{readout} used for label prediction.

The objective is to minimize the loss between the predicted label and true label.



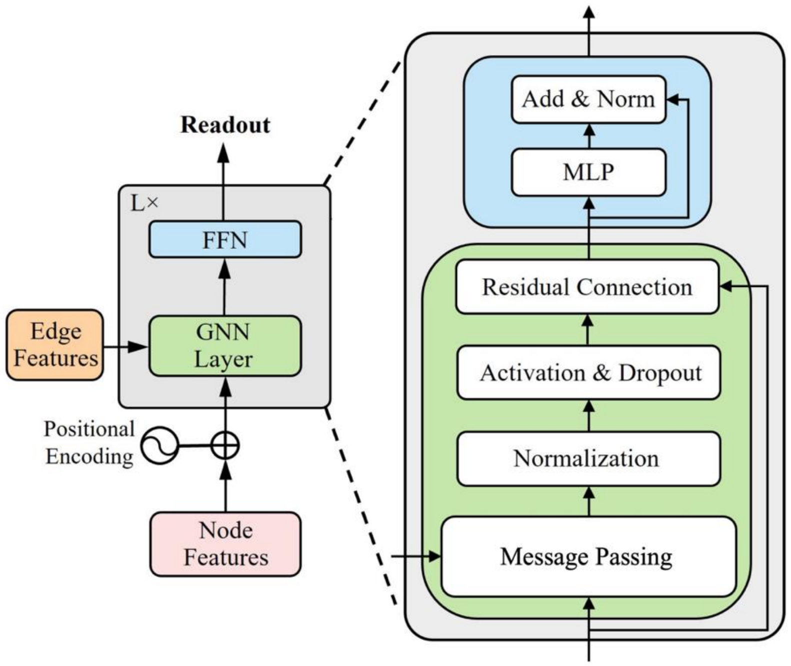
GNN+: Enhanced Classic GNNs Framework

Message-Passing GNNs



$$\boldsymbol{h}_v^l = \sigma\Big(\sum_{u\in\mathcal{N}(v)\cup\{v\}} \frac{1}{\sqrt{\hat{d}_u \hat{d}_v}} \boldsymbol{h}_u^{l-1} \boldsymbol{W}^l\Big)$$

GNN+ Framework

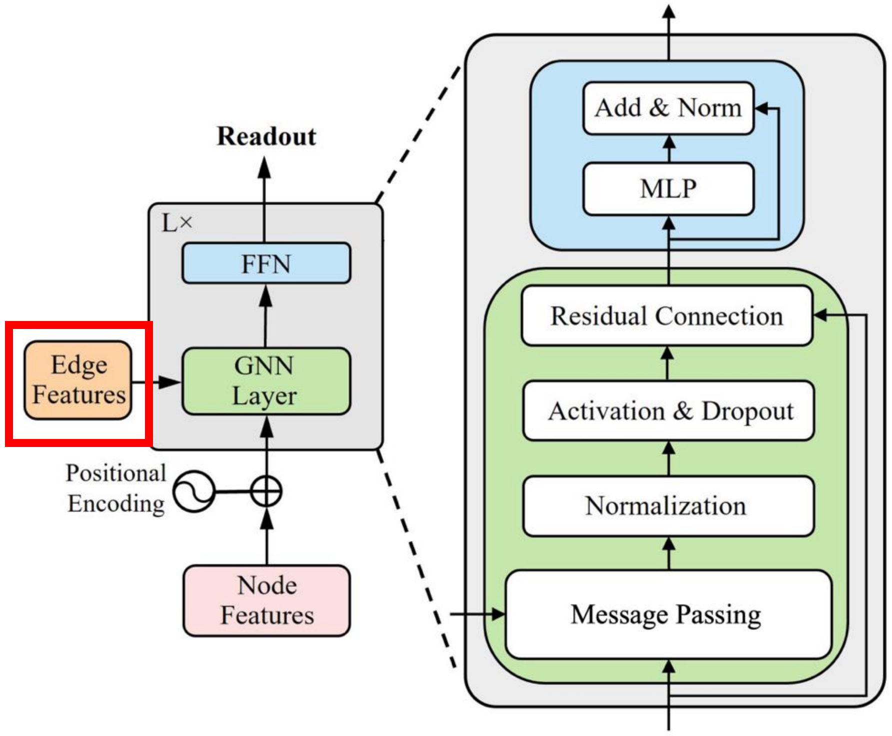


Proposed the GNN+ architecture, which integrates 6 existing hyperparameter techniques into the message-passing mechanism.

GNN+: Enhanced Classic GNNs Framework

Edge features were initially incorporated into some classic GNN frameworks by directly integrating them into the message-passing process to enhance information propagation between nodes. Taking GCN as an example:

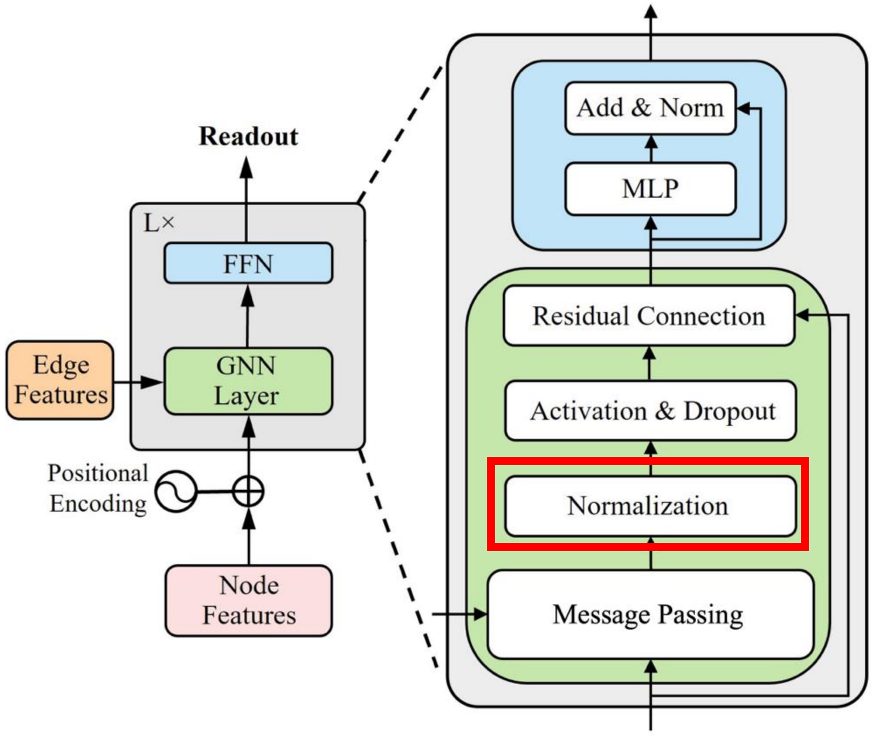
$$\boldsymbol{h}_v^l = \sigma\Big(\sum_{u \in \mathcal{N}(v)\cup\{v\}} \frac{1}{\sqrt{\hat{d}_u \hat{d}_v}} \boldsymbol{h}_u^{l-1}\boldsymbol{W}^l + \boldsymbol{e}_{uv}\boldsymbol{W}_e^l\Big),$$



GNN+: Enhanced Classic GNNs Framework

Batch Normalization (BN) and Layer Normalization (LN) are widely used techniques, typically applied to the output of each layer before the activation function. Here, we use BN:

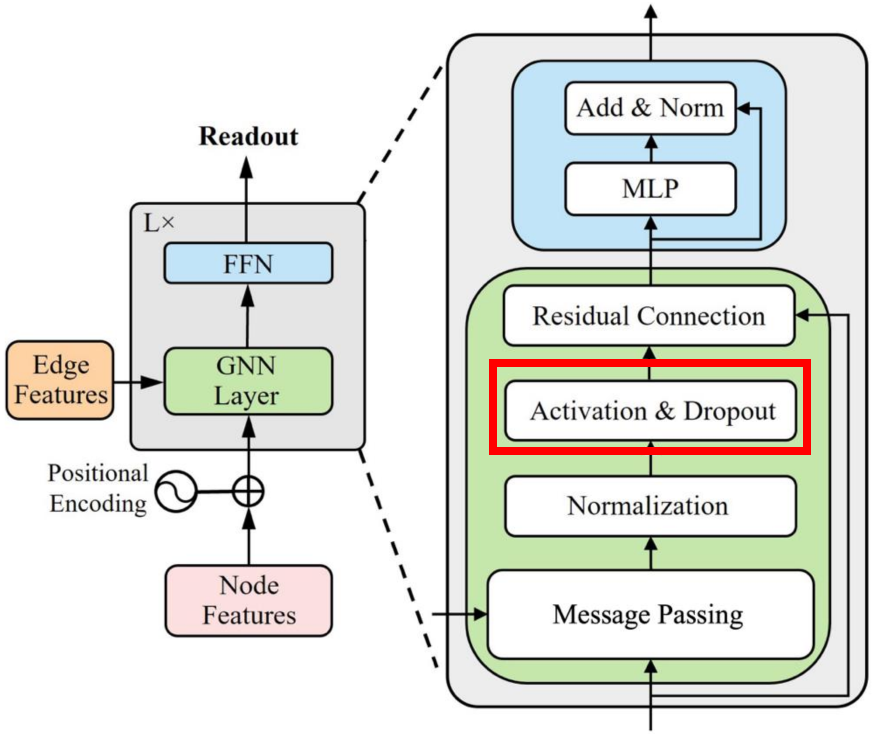
$$\boldsymbol{h}_v^l = \sigma(\mathrm{BN}(\sum_{u\in\mathcal{N}(v)\cup\{v\}} \frac{1}{\sqrt{\hat{d}_u\hat{d}_v}}\boldsymbol{h}_u^{l-1}\boldsymbol{W}^l + \boldsymbol{e}_{uv}\boldsymbol{W}_e^l)).$$



GNN+: Enhanced Classic GNNs Framework

Dropout is applied to the embeddings after activation:

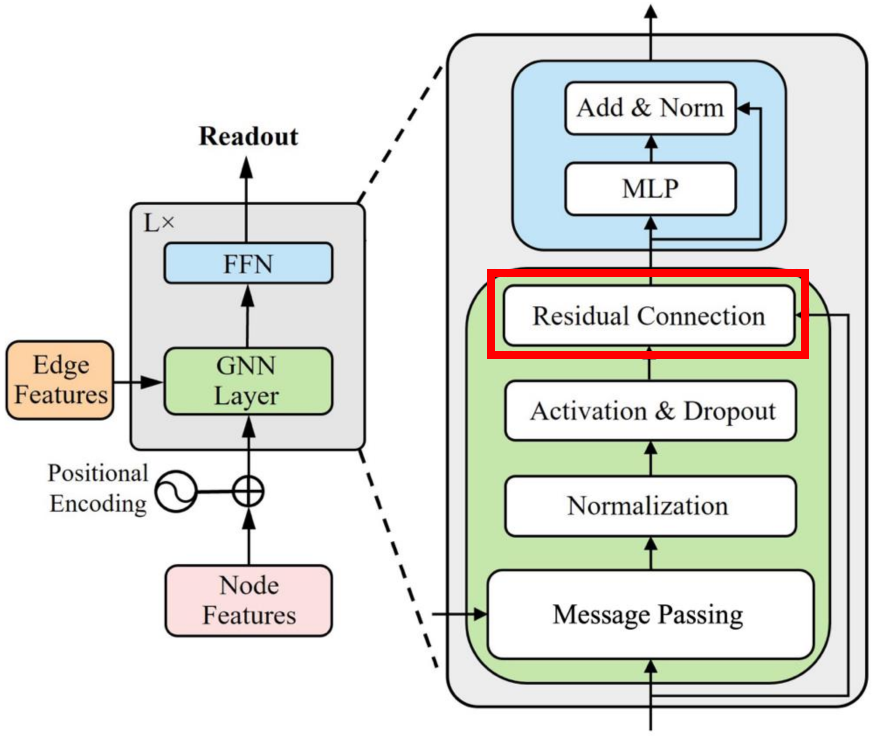
$$\boldsymbol{h}_v^l = \text{Dropout}(\sigma(\text{BN}(\sum_{u \in \mathcal{N}(v) \cup \{v\}} \frac{1}{\sqrt{\hat{d}_u \hat{d}_v}} \boldsymbol{h}_u^{l-1} \boldsymbol{W}^l + \boldsymbol{e}_{uv} \boldsymbol{W}_e^l))).$$



GNN+: Enhanced Classic GNNs Framework

Residual connections can be integrated into GNNs as follows:

$$\boldsymbol{h}_v^l = \text{Dropout}(\sigma(\text{BN}(\sum_{u \in \mathcal{N}(v) \cup \{v\}} \frac{1}{\sqrt{\hat{d}_u \hat{d}_v}} \boldsymbol{h}_u^{l-1} \boldsymbol{W}^l + \boldsymbol{e}_{uv} \boldsymbol{W}_e^l))) + \boxed{\boldsymbol{h}_v^{l-1}}.$$



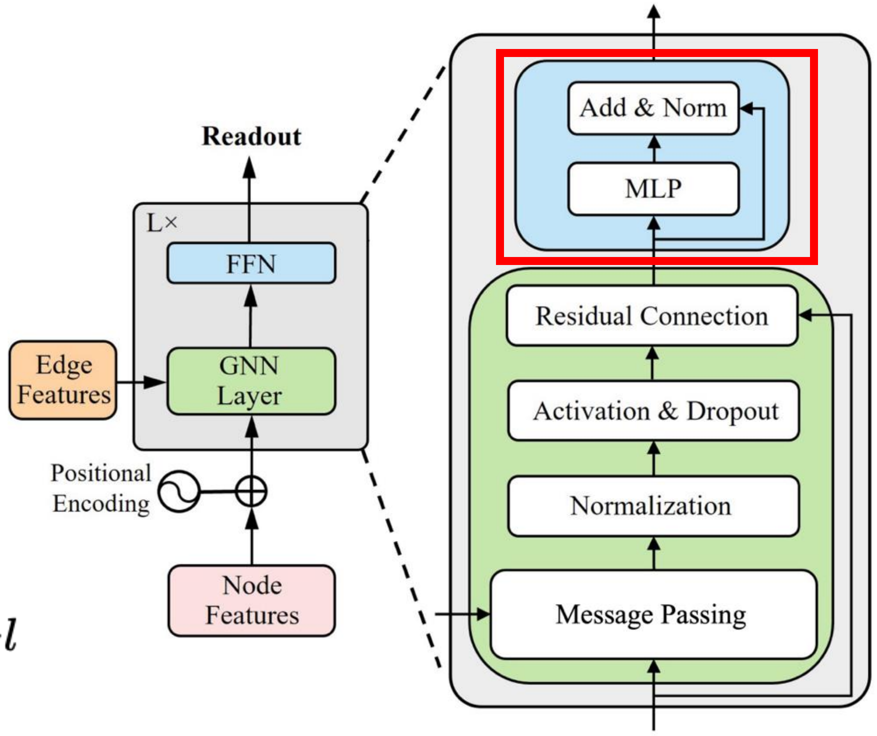
GNN+: Enhanced Classic GNNs Framework

Transformer incorporate a feed-forward network (FFN) as a crucial component within each of their layers.

Inspired by this, we propose appending a fully-connected FFN at the end of each layer of GNNs, defined as:

$$\boxed{\mathrm{FFN}(\boldsymbol{h}) = \mathrm{BN}(\sigma(\boldsymbol{h}\boldsymbol{W}^{l}_{\mathrm{FFN}_1})\boldsymbol{W}^{l}_{\mathrm{FFN}_2} + \boldsymbol{h}),}$$

$$\boldsymbol{h}^l_v = \mathrm{FFN}(\mathrm{Dropout}(\sigma(\mathrm{BN}(\sum_{u \in \mathcal{N}(v)\cup\{v\}} \frac{1}{\sqrt{\hat{d}_u \hat{d}_v}} \boldsymbol{h}^{l-1}_u \boldsymbol{W}^l + \boldsymbol{e}_{uv}\boldsymbol{W}^l_e))) + \boldsymbol{h}^{l-1}_v).$$

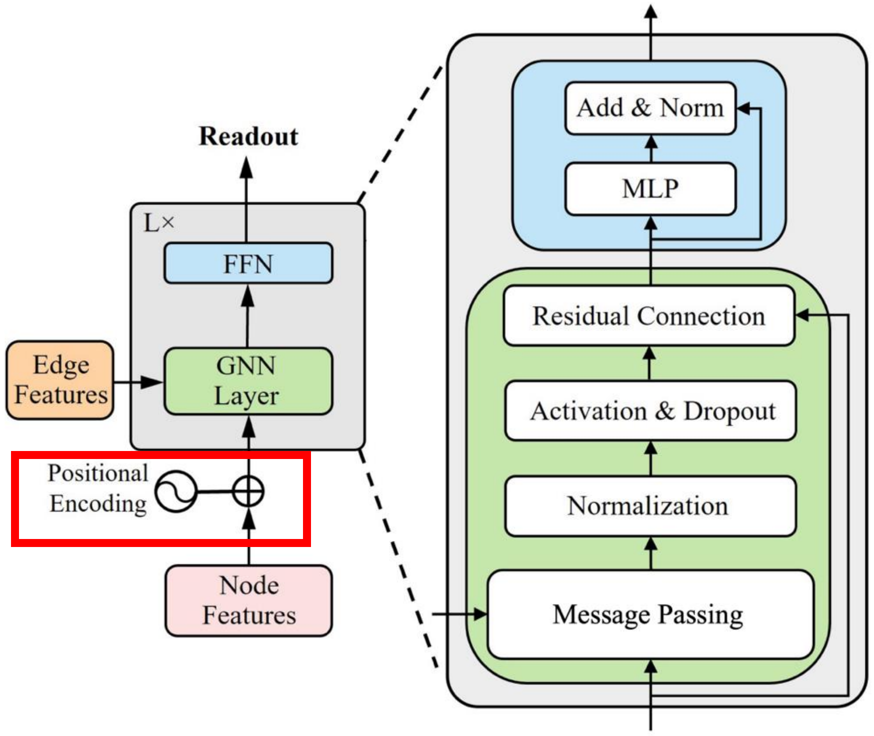


GNN+: Enhanced Classic GNNs Framework

Positional encoding (PE) was introduced in the Transformer to represent the positions of tokens within a sequence for language modeling. Various PE methods have been proposed for graph, such as LapPE, RWSE.

Following the practice, we use RWSE to improve the performance of GNNs as follows:

$$\boldsymbol{x}_v = [\boldsymbol{x}_v \| \boldsymbol{x}_v^{\text{RWSE}}] \boldsymbol{W}_{\text{PE}},$$



Graph-level Datasets

Table 7. Overview of the datasets used for graph-level tasks (Dwivedi et al., 2023; 2022; Hu et al., 2020; Freitas & Dong, 2021).

Dataset	# graphs	Avg. # nodes	Avg. # edges	# node/edge feats	Prediction level	Prediction task	Metric
ZINC	12,000	23.2	24.9	28/1	graph	regression	MAE
MNIST	70,000	70.6	564.5	3/1	graph	10-class classif.	Accuracy
CIFAR10	60,000	117.6	941.1	5/1	graph	10-class classif.	Accuracy
PATTERN	14,000	118.9	3,039.3	3/1	inductive node	binary classif.	Accuracy
CLUSTER	12,000	117.2	2,150.9	7/1	inductive node	6-class classif.	Accuracy
Peptides-func	15,535	150.9	307.3	9/3	graph	10-task classif.	Avg. Precision
Peptides-struct	15,535	150.9	307.3	9/3	graph	11-task regression	MAE
PascalVOC-SP	11,355	479.4	2,710.5	14/2	inductive node	21-class classif.	F1 score
COCO-SP	123,286	476.9	2,693.7	14/2	inductive node	81-class classif.	F1 score
MalNet-Tiny	5,000	1,410.3	2,859.9	5/1	graph	5-class classif.	Accuracy
ogbg-molhiv	41,127	25.5	27.5	9/3	graph	binary classif.	AUROC
ogbg-molpcba	437,929	26.0	28.1	9/3	graph	128-task classif.	Avg. Precision
ogbg-ppa	158,100	243.4	2,266.1	1/7	graph	37-task classif.	Accuracy
ogbg-code2	452,741	125.2	124.2	2/2	graph	5 token sequence	F1 score

3 classic GNNs for graph-level tasks: GCN, GIN, GatedGCN

Empirical Findings

Table 2. Test performance on five benchmarks from (Dwivedi et al., 2023) (%). Shown is the mean ± s.d. of 5 runs with different random seeds. $^+$ denotes the enhanced version, while the baseline results were obtained from their respective original papers. # Param ∼ 500K for ZINC, PATTERN, and CLUSTER, and ∼ 100K for MNIST and CIFAR10. The top 1^{st}, 2^{nd} and 3^{rd} results are highlighted.

	ZINC	MNIST	CIFAR10	PATTERN	CLUSTER
# graphs	12,000	70,000	60,000	14,000	12,000
Avg. # nodes	23.2	70.6	117.6	118.9	117.2
Avg. # edges	24.9	564.5	941.1	3039.3	2150.9
Metric	MAE ↓	Accuracy ↑	Accuracy ↑	Accuracy ↑	Accuracy ↑
GT (2020)	0.226 ± 0.014	90.831 ± 0.161	59.753 ± 0.293	84.808 ± 0.068	73.169 ± 0.622
SAN (2021)	0.139 ± 0.006	–	–	86.581 ± 0.037	76.691 ± 0.650
Graphormer (2021)	0.122 ± 0.006	–	–	–	–
SAT (2022)	0.094 ± 0.008	–	–	86.848 ± 0.037	77.856 ± 0.104
EGT (2022)	0.108 ± 0.009	98.173 ± 0.087	68.702 ± 0.409	86.821 ± 0.020	79.232 ± 0.348
GraphGPS (2022)	0.070 ± 0.004	98.051 ± 0.126	72.298 ± 0.356	86.685 ± 0.059	78.016 ± 0.180
GRPE (2022)	0.094 ± 0.002	–	–	87.020 ± 0.042	–
Graphormer-URPE (2022)	0.086 ± 0.007	–	–	–	–
Graphormer-GD (2023)	0.081 ± 0.009	–	–	–	–
Specformer (2023)	0.066 ± 0.003	–	–	–	–
LGI-GT (2023)	–	–	–	86.930 ± 0.040	–
GPTrans-Nano (2023b)	–	–	–	86.731 ± 0.085	–
Graph ViT/MLP-Mixer (2023)	0.073 ± 0.001	98.460 ± 0.090	73.960 ± 0.330	–	–
Exphormer (2023)	–	98.414 ± 0.038	74.754 ± 0.194	86.734 ± 0.008	–
GRIT (2023)	0.059 ± 0.002	98.108 ± 0.111	76.468 ± 0.881	87.196 ± 0.076	80.026 ± 0.277
GRED (2024)	0.077 ± 0.002	98.383 ± 0.012	76.853 ± 0.185	86.759 ± 0.020	78.495 ± 0.103
GEAET (2024)	–	98.513 ± 0.086	76.634 ± 0.427	86.993 ± 0.026	–
TIGT (2024)	0.057 ± 0.002	98.231 ± 0.132	73.963 ± 0.361	86.681 ± 0.062	78.025 ± 0.223
Cluster-GT (2024a)	0.071 ± 0.004	–	–	–	–
GMN (2024)	–	98.391 ± 0.182	74.560 ± 0.381	87.090 ± 1.260	–
Graph-Mamba (2024)	–	98.420 ± 0.080	73.700 ± 0.340	86.710 ± 0.050	76.800 ± 0.360
GCN	0.367 ± 0.011	90.705 ± 0.218	55.710 ± 0.381	71.892 ± 0.334	68.498 ± 0.976
GCN$^+$	0.076 ± 0.009 **79.3%↓**	98.382 ± 0.095 **8.5%↑**	69.824 ± 0.413 **25.4%↑**	87.021 ± 0.095 **21.1%↑**	77.109 ± 0.872 **12.6%↑**
GIN	0.526 ± 0.051	96.485 ± 0.252	55.255 ± 1.527	85.387 ± 0.136	64.716 ± 1.553
GIN$^+$	0.065 ± 0.004 **87.6%↓**	98.285 ± 0.103 **1.9%↑**	69.592 ± 0.287 **25.9%↑**	86.842 ± 0.048 **1.7%↑**	74.794 ± 0.213 **15.6%↑**
GatedGCN	0.282 ± 0.015	97.340 ± 0.143	67.312 ± 0.311	85.568 ± 0.088	73.840 ± 0.326
GatedGCN$^+$	0.077 ± 0.005 **72.7%↓**	98.712 ± 0.137 **1.4%↑**	77.218 ± 0.381 **14.7%↑**	87.029 ± 0.037 **1.7%↑**	79.128 ± 0.235 **7.1%↑**
Time (epoch) of GraphGPS	21s	76s	64s	32s	86s
Time (epoch) of **GCN$^+$**	**7s**	**60s**	**40s**	**19s**	**29s**

Empirical Findings

Table 3. Test performance on five datasets from Long-Range Graph Benchmarks (LRGB) (Dwivedi et al., 2022; Freitas & Dong, 2021). $^+$ denotes the enhanced version, while the baseline results were obtained from their respective original papers. # Param ∼ 500K for all.

	Peptides-func	Peptides-struct	PascalVOC-SP	COCO-SP	MalNet-Tiny
# graphs	15,535	15,535	11,355	123,286	5,000
Avg. # nodes	150.9	150.9	479.4	476.9	1,410.3
Avg. # edges	307.3	307.3	2,710.5	2,693.7	2,859.9
Metric	Avg. Precision ↑	MAE ↓	F1 score ↑	F1 score ↑	Accuracy ↑
GT (2020)	0.6326 ± 0.0126	0.2529 ± 0.0016	0.2694 ± 0.0098	0.2618 ± 0.0031	–
SAN (2021)	0.6439 ± 0.0075	0.2545 ± 0.0012	0.3230 ± 0.0039	0.2592 ± 0.0158	–
GraphGPS (2022)	0.6535 ± 0.0041	0.2500 ± 0.0005	0.3748 ± 0.0109	0.3412 ± 0.0044	0.9350 ± 0.0041
GraphGPS (2023)	0.6534 ± 0.0091	0.2509 ± 0.0014	0.4440 ± 0.0065	0.3884 ± 0.0055	0.9350 ± 0.0041
NAGphormer (2023a)	–	–	0.4006 ± 0.0061	0.3458 ± 0.0070	–
DIFFormer (2023)	–	–	0.3988 ± 0.0045	0.3620 ± 0.0012	–
MGT (2023)	0.6817 ± 0.0064	0.2453 ± 0.0025	–	–	–
DRew (2023)	0.7150 ± 0.0044	0.2536 ± 0.0015	0.3314 ± 0.0024	–	–
Graph ViT/MLP-Mixer (2023)	0.6970 ± 0.0080	0.2449 ± 0.0016	–	–	–
Exphormer (2023)	0.6258 ± 0.0092	0.2512 ± 0.0025	0.3446 ± 0.0064	0.3430 ± 0.0108	0.9402 ± 0.0021
GRIT (2023)	0.6988 ± 0.0082	0.2460 ± 0.0012	–	–	–
Subgraphormer (2024)	0.6415 ± 0.0052	0.2475 ± 0.0007	–	–	–
GRED (2024)	0.7133 ± 0.0011	0.2455 ± 0.0013	–	–	–
GEAET (2024)	0.6485 ± 0.0035	0.2547 ± 0.0009	0.3933 ± 0.0027	0.3219 ± 0.0052	–
TIGT (2024)	0.6679 ± 0.0074	0.2485 ± 0.0015	–	–	–
GECO (2024)	0.6975 ± 0.0025	0.2464 ± 0.0009	0.4210 ± 0.0080	0.3320 ± 0.0032	–
GPNN (2024)	0.6955 ± 0.0057	0.2454 ± 0.0003	–	–	–
Graph-Mamba (2024)	0.6739 ± 0.0087	0.2478 ± 0.0016	0.4191 ± 0.0126	0.3960 ± 0.0175	0.9340 ± 0.0027
GSSC (2024b)	0.7081 ± 0.0062	0.2459 ± 0.0020	0.4561 ± 0.0039	–	0.9406 ± 0.0064
GCN	0.6860 ± 0.0050	0.2460 ± 0.0007	0.2078 ± 0.0031	0.1338 ± 0.0007	0.8100 ± 0.0081
GCN$^+$	0.7261 ± 0.0067 **5.9%↑**	0.2421 ± 0.0016 **1.6%↓**	0.3357 ± 0.0087 **62.0%↑**	0.2733 ± 0.0041 **104.9%↑**	0.9354 ± 0.0045 **15.5%↑**
GIN	0.6621 ± 0.0067	0.2473 ± 0.0017	0.2718 ± 0.0054	0.2125 ± 0.0009	0.8898 ± 0.0055
GIN$^+$	0.7059 ± 0.0089 **6.6%↑**	0.2429 ± 0.0019 **1.8%↓**	0.3189 ± 0.0105 **17.3%↑**	0.2483 ± 0.0046 **16.9%↑**	0.9325 ± 0.0040 **4.8%↑**
GatedGCN	0.6765 ± 0.0047	0.2477 ± 0.0009	0.3880 ± 0.0040	0.2922 ± 0.0018	0.9223 ± 0.0065
GatedGCN$^+$	0.7006 ± 0.0033 **3.6%↑**	0.2431 ± 0.0020 **1.9%↓**	0.4263 ± 0.0057 **9.9%↑**	0.3802 ± 0.0015 **30.1%↑**	0.9460 ± 0.0057 **2.6%↑**
Time (epoch) of GraphGPS	6s	6s	17s	213s	46s
Time (epoch) of **GCN$^+$**	6s	6s	**12s**	**162s**	**6s**

Empirical Findings

Table 4. Test performance in four benchmarks from Open Graph Benchmark (OGB) (Hu et al., 2020). $^{+}$ denotes the enhanced version, while the baseline results were obtained from their respective original papers. † indicates the use of additional pretraining datasets, included here for reference only and excluded from ranking.

	ogbg-molhiv	ogbg-molpcba	ogbg-ppa	ogbg-code2
# graphs	41,127	437,929	158,100	452,741
Avg. # nodes	25.5	26.0	243.4	125.2
Avg. # edges	27.5	28.1	2,266.1	124.2
Metric	AUROC ↑	Avg. Precision ↑	Accuracy ↑	F1 score ↑
GT (2020)	–	–	0.6454 ± 0.0033	0.1670 ± 0.0015
GraphTrans (2021)	–	0.2761 ± 0.0029	–	0.1830 ± 0.0024
SAN (2021)	0.7785 ± 0.2470	0.2765 ± 0.0042	–	–
Graphormer (pre-trained) (2021)	0.8051 ± 0.0053†	–	–	–
SAT (2022)	–	–	0.7522 ± 0.0056	**0.1937 ± 0.0028**
EGT (pre-trained) (2022)	0.8060 ± 0.0065†	0.2961 ± 0.0024†	–	–
GraphGPS (2022)	0.7880 ± 0.0101	0.2907 ± 0.0028	0.8015 ± 0.0033	0.1894 ± 0.0024
Specformer (2023)	0.7889 ± 0.0124	**0.2972 ± 0.0023**	–	–
Graph ViT/MLP-Mixer (2023)	0.7997 ± 0.0102	–	–	–
Exphormer (2023)	0.7834 ± 0.0044	0.2849 ± 0.0025	–	–
GRIT (2023)	0.7835 ± 0.0054	0.2362 ± 0.0020	–	–
Subgraphormer (2024)	**0.8038 ± 0.0192**	–	–	–
GECO (2024)	0.7980 ± 0.0200	**0.2961 ± 0.0008**	0.7982 ± 0.0042	**0.1915 ± 0.0020**
GSSC (2024b)	**0.8035 ± 0.0142**	–	–	–
GCN	0.7606 ± 0.0097	0.2020 ± 0.0024	0.6839 ± 0.0084	0.1507 ± 0.0018
GCN^{+}	0.8012 ± 0.0124 **5.4%↑**	0.2721 ± 0.0046 **34.7%↑**	**0.8077 ± 0.0041** **18.1%↑**	0.1787 ± 0.0026 **18.6%↑**
GIN	0.7835 ± 0.0125	0.2266 ± 0.0028	0.6892 ± 0.0100	0.1495 ± 0.0023
GIN^{+}	0.7928 ± 0.0099 **1.2%↑**	0.2703 ± 0.0024 **19.3%↑**	**0.8107 ± 0.0053** **17.7%↑**	0.1803 ± 0.0019 **20.6%↑**
GatedGCN	0.7687 ± 0.0136	0.2670 ± 0.0020	0.7531 ± 0.0083	0.1606 ± 0.0015
GatedGCN^{+}	**0.8040 ± 0.0164** **4.6%↑**	**0.2981 ± 0.0024** **11.6%↑**	**0.8258 ± 0.0055** **9.7%↑**	**0.1896 ± 0.0024** **18.1%↑**
Time (epoch/s) of GraphGPS	96s	196s	276s	1919s
Time (epoch/s) of **GCN^{+}**	**16s**	**91s**	**178s**	**476s**

Empirical Findings

The enhanced versions of classic GNNs achieved state-of-the-art performance, ranking in the top three across 14 datasets, including first place in 8 of them, while also demonstrating superior efficiency.

Table 4. Test performance in four benchmarks from Open Graph Benchmark (OGB) (Hu et al., 2020). $^{+}$ denotes the enhanced version, while the baseline results were obtained from their respective original papers. † indicates the use of additional pretraining datasets, included here for reference only and excluded from ranking.

	ogbg-molhiv	ogbg-molpcba	ogbg-ppa	ogbg-code2
# graphs	41,127	437,929	158,100	452,741
Avg. # nodes	25.5	26.0	243.4	125.2
Avg. # edges	27.5	28.1	2,266.1	124.2
Metric	AUROC ↑	Avg. Precision ↑	Accuracy ↑	F1 score ↑
GT (2020)	–	–	0.6454 ± 0.0033	0.1670 ± 0.0015
GraphTrans (2021)	–	0.2761 ± 0.0029	–	0.1830 ± 0.0024
SAN (2021)	0.7785 ± 0.2470	0.2765 ± 0.0042	–	–
Graphormer (pre-trained) (2021)	$0.8051 \pm 0.0053^{\dagger}$	–	–	–
SAT (2022)	–	–	0.7522 ± 0.0056	0.1937 ± 0.0028
EGT (pre-trained) (2022)	$0.8060 \pm 0.0065^{\dagger}$	$0.2961 \pm 0.0024^{\dagger}$	–	–
GraphGPS (2022)	0.7880 ± 0.0101	0.2907 ± 0.0028	0.8015 ± 0.0033	0.1894 ± 0.0024
Specformer (2023)	0.7889 ± 0.0124	0.2972 ± 0.0023	–	–
Graph ViT/MLP-Mixer (2023)	0.7997 ± 0.0102	–	–	–
Exphormer (2023)	0.7834 ± 0.0044	0.2849 ± 0.0025	–	–
GRIT (2023)	0.7835 ± 0.0054	0.2362 ± 0.0020	–	–
Subgraphormer (2024)	0.8038 ± 0.0192	–	–	–
GECO (2024)	0.7980 ± 0.0200	0.2961 ± 0.0008	0.7982 ± 0.0042	0.1915 ± 0.0020
GSSC (2024b)	0.8035 ± 0.0142	–	–	–
GCN	0.7606 ± 0.0097	0.2020 ± 0.0024	0.6839 ± 0.0084	0.1507 ± 0.0018
GCN^{+}	0.8012 ± 0.0124 **5.4%↑**	0.2721 ± 0.0046 **34.7%↑**	0.8077 ± 0.0041 **18.1%↑**	0.1787 ± 0.0026 **18.6%↑**
GIN	0.7835 ± 0.0125	0.2266 ± 0.0028	0.6892 ± 0.0100	0.1495 ± 0.0023
GIN^{+}	0.7928 ± 0.0099 **1.2%↑**	0.2703 ± 0.0024 **19.3%↑**	0.8107 ± 0.0053 **17.7%↑**	0.1803 ± 0.0019 **20.6%↑**
GatedGCN	0.7687 ± 0.0136	0.2670 ± 0.0020	0.7531 ± 0.0083	0.1606 ± 0.0015
GatedGCN^{+}	0.8040 ± 0.0164 **4.6%↑**	0.2981 ± 0.0024 **11.6%↑**	0.8258 ± 0.0055 **9.7%↑**	0.1896 ± 0.0024 **18.1%↑**
Time (epoch/s) of GraphGPS	96s	196s	276s	1919s
Time (epoch/s) of **GCN^{+}**	**16s**	**91s**	**178s**	**476s**

Ablation Studies - Edge Features

- **The integration of edge features is particularly effective in molecular and image superpixel datasets, where these features carry critical information.**

Table 6. Ablation study on LRGB and OGB datasets. - indicates that the corresponding hyperparameter is not used in GNN^{+}, as it empirically leads to inferior performance.

Metric	Peptides-func Avg. Precision ↑	Peptides-struct MAE ↓	PascalVOC-SP F1 score ↑	COCO-SP F1 score ↑	MalNet-Tiny Accuracy ↑	ogbg-molhiv AUROC ↑	ogbg-molpcba Avg. Precision ↑	ogbg-ppa Accuracy ↑	ogbg-code2 F1 score ↑
GCN^{+}	**0.7261** ± 0.0067	**0.2421** ± 0.0016	**0.3357** ± 0.0087	**0.2733** ± 0.0041	**0.9354** ± 0.0045	**0.8012** ± 0.0124	**0.2721** ± 0.0046	**0.8077** ± 0.0041	**0.1787** ± 0.0026
(-) Edge.	0.7191 ± 0.0036	–	0.2942 ± 0.0043	0.2219 ± 0.0060	0.9292 ± 0.0034	0.7714 ± 0.0204	0.2628 ± 0.0019	0.2994 ± 0.0062	0.1785 ± 0.0033
(-) Norm	0.7107 ± 0.0027	0.2509 ± 0.0026	0.1802 ± 0.0111	0.2332 ± 0.0079	0.9236 ± 0.0054	0.7753 ± 0.0049	0.2528 ± 0.0016	0.6705 ± 0.0104	0.1679 ± 0.0027
(-) Dropout	0.6748 ± 0.0055	0.2549 ± 0.0025	0.3072 ± 0.0069	0.2601 ± 0.0046	–	0.7431 ± 0.0185	0.2405 ± 0.0047	0.7893 ± 0.0052	0.1641 ± 0.0043
(-) RC	–	–	0.2734 ± 0.0036	0.1948 ± 0.0096	0.8916 ± 0.0048	–	–	0.7520 ± 0.0157	0.1785 ± 0.0029
(-) FFN	–	–	0.2786 ± 0.0068	0.2314 ± 0.0073	0.9118 ± 0.0078	0.7432 ± 0.0052	0.2621 ± 0.0019	0.7672 ± 0.0071	0.1594 ± 0.0020
(-) PE	0.7069 ± 0.0093	0.2447 ± 0.0015	–	–	–	0.7593 ± 0.0051	0.2667 ± 0.0034	–	–
GIN^{+}	**0.7059** ± 0.0089	**0.2429** ± 0.0019	**0.3189** ± 0.0105	**0.2483** ± 0.00[illegible]6	**0.9325** ± 0.0040	**0.7928** ± 0.0099	**0.2703** ± 0.0024	**0.8107** ± 0.0053	**0.1803** ± 0.0019
(-) Edge.	0.7033 ± 0.0015	0.2442 ± 0.0028	0.2956 ± 0.0047	0.2259 ± 0.00[illegible]3	0.9286 ± 0.0049	0.7597 ± 0.0103	0.2702 ± 0.0021	0.2789 ± 0.0031	0.1752 ± 0.0020
(-) Norm	0.6934 ± 0.0077	0.2444 ± 0.0015	0.2707 ± 0.0037	0.2244 ± 0.0063	0.9322 ± 0.0025	0.7874 ± 0.0114	0.2556 ± 0.0026	0.6484 ± 0.0246	0.1722 ± 0.0034
(-) Dropout	0.6384 ± 0.0094	0.2531 ± 0.0030	0.3153 ± 0.0113	–	–	–	0.2545 ± 0.0068	0.7673 ± 0.0059	0.1730 ± 0.0018
(-) RC	0.6975 ± 0.0038	0.2527 ± 0.0015	0.2350 ± 0.0044	0.1741 ± 0.0085	0.9150 ± 0.0047	0.7733 ± 0.0122	0.1454 ± 0.0061	–	0.1617 ± 0.0026
(-) FFN	–	–	0.2393 ± 0.0049	0.1599 ± 0.0081	0.8944 ± 0.0074	–	0.2534 ± 0.0033	0.6676 ± 0.0039	0.1491 ± 0.0016
(-) PE	0.6855 ± 0.0027	0.2455 ± 0.0019	0.3141 ± 0.0031	–	–	0.7791 ± 0.0268	0.2601 ± 0.0023	–	–
GatedGCN^{+}	**0.7006** ± 0.0033	**0.2431** ± 0.0020	**0.4263** ± 0.0057	**0.3802** ± 0.0015	**0.9460** ± 0.0057	**0.8040** ± 0.0164	**0.2981** ± 0.0024	**0.8258** ± 0.0055	**0.1896** ± 0.0024
(-) Edge.	0.6882 ± 0.0028	0.2466 ± 0.0018	0.3764 ± 0.0117	0.3172 ± 0.0109	0.9372 ± 0.0062	0.7831 ± 0.0157	0.2951 ± 0.0028	0.0948 ± 0.0000	0.1891 ± 0.0021
(-) Norm	0.6733 ± 0.0026	0.2474 ± 0.0015	0.3628 ± 0.0043	0.3527 ± 0.0051	0.9326 ± 0.0056	0.7879 ± 0.0178	0.2748 ± 0.0012	0.6864 ± 0.0165	0.1743 ± 0.0026
(-) Dropout	0.6695 ± 0.0101	0.2508 ± 0.0014	0.3389 ± 0.0066	0.3393 ± 0.0051	–	–	0.2582 ± 0.0036	0.8088 ± 0.0062	0.1724 ± 0.0027
(-) RC	–	0.2498 ± 0.0034	0.4075 ± 0.0052	0.3475 ± 0.0064	0.9402 ± 0.0054	0.7833 ± 0.0177	0.2897 ± 0.0016	0.8099 ± 0.0053	0.1844 ± 0.0025
(-) FFN	–	–	–	0.3508 ± 0.0049	0.9364 ± 0.0059	–	0.2875 ± 0.0022	–	0.1718 ± 0.0024
(-) PE	0.6729 ± 0.0084	0.2461 ± 0.0025	0.4052 ± 0.0031	–	–	0.7771 ± 0.0057	0.2813 ± 0.0022	–	–

Ablation Studies - Normalization

- **Normalization tends to have a greater impact on larger-scale datasets, whereas its impact is less significant on smaller datasets.**

Table 5. Ablation study on GNN Benchmark (Dwivedi et al., 2023) (%). - indicates that the corresponding hyperparameter is not used in GNN$^+$, as it empirically leads to inferior performance.

Metric	ZINC MAE ↓	MNIST Accuracy ↑	CIFAR10 Accuracy ↑	PATTERN Accuracy ↑	CLUSTER Accuracy ↑
GCN$^+$	**0.076** ± 0.009	**98.382** ± 0.095	**69.824** ± 0.413	**87.021** ± 0.095	**77.109** ± 0.872
(-) Edge.	0.135 ± 0.004	98.153 ± 0.042	68.256 ± 0.357	86.854 ± 0.054	–
(-) Norm	0.107 ± 0.011	97.886 ± 0.066	60.765 ± 0.829	52.769 ± 0.874	16.563 ± 0.134
(-) Dropout	–	97.897 ± 0.071	65.693 ± 0.461	86.764 ± 0.045	74.926 ± 0.469
(-) RC	0.159 ± 0.016	95.929 ± 0.169	58.186 ± 0.295	86.059 ± 0.274	16.508 ± 0.615
(-) FFN	0.132 ± 0.021	97.174 ± 0.063	63.573 ± 0.346	86.746 ± 0.088	72.606 ± 1.243
(-) PE	0.127 ± 0.010	–	–	85.597 ± 0.241	75.568 ± 1.147
GIN$^+$	**0.065** ± 0.004	**98.285** ± 0.103	**69.592** ± 0.287	**86.842** ± 0.048	**74.794** ± 0.213
(-) Edge.	0.122 ± 0.009	97.655 ± 0.075	68.196 ± 0.107	86.714 ± 0.036	65.895 ± 3.425
(-) Norm	0.096 ± 0.006	97.695 ± 0.065	64.918 ± 0.059	86.815 ± 0.855	72.119 ± 0.359
(-) Dropout	–	98.214 ± 0.064	66.638 ± 0.873	86.836 ± 0.053	73.316 ± 0.355
(-) RC	0.137 ± 0.031	97.675 ± 0.175	64.910 ± 0.102	86.645 ± 0.125	16.800 ± 0.088
(-) FFN	0.104 ± 0.003	11.350 ± 0.008	60.582 ± 0.395	58.511 ± 0.016	62.175 ± 2.895
(-) PE	0.123 ± 0.014	–	–	86.592 ± 0.049	73.925 ± 0.165
GatedGCN$^+$	**0.077** ± 0.005	**98.712** ± 0.137	**77.218** ± 0.381	**87.029** ± 0.037	**79.128** ± 0.235
(-) Edge.	0.119 ± 0.001	98.085 ± 0.045	72.128 ± 0.275	86.879 ± 0.017	76.075 ± 0.845
(-) Norm	0.088 ± 0.003	98.275 ± 0.045	71.995 ± 0.445	86.942 ± 0.023	78.495 ± 0.155
(-) Dropout	0.089 ± 0.003	98.225 ± 0.095	70.383 ± 0.429	86.802 ± 0.034	77.597 ± 0.126
(-) RC	0.106 ± 0.002	98.442 ± 0.067	75.149 ± 0.155	86.845 ± 0.025	16.670 ± 0.307
(-) FFN	0.098 ± 0.005	98.438 ± 0.151	76.243 ± 0.131	86.935 ± 0.025	78.975 ± 0.145
(-) PE	0.174 ± 0.009	–	–	85.595 ± 0.065	77.515 ± 0.265

Table 6. Ablation study on LRGB and OGB datasets. - indicates that the corresponding hyperparameter is not used in GNN$^+$, as it empirically leads to inferior performance.

Metric	Peptides-func Avg. Precision ↑	Peptides-struct MAE ↓	PascalVOC-SP F1 score ↑	COCO-SP F1 score ↑	MalNet-Tiny Accuracy ↑	ogbg-molhiv AUROC ↑	ogbg-molpcba Avg. Precision ↑	ogbg-ppa Accuracy ↑	ogbg-code2 F1 score ↑
GCN$^+$	**0.7261** ± 0.0067	**0.2421** ± 0.0016	**0.3357** ± 0.0087	**0.2733** ± 0.0041	**0.9354** ± 0.0045	**0.8012** ± 0.0124	**0.2721** ± 0.0046	**0.8077** ± 0.0041	**0.1787** ± 0.0026
(-) Edge.	0.7191 ± 0.0036	–	0.2942 ± 0.0043	0.2219 ± 0.0060	0.9292 ± 0.0034	0.7714 ± 0.0204	0.2628 ± 0.0019	0.2994 ± 0.0062	0.1785 ± 0.0033
(-) Norm	0.7107 ± 0.0027	0.2509 ± 0.0026	0.1802 ± 0.0111	0.2332 ± 0.0079	0.9236 ± 0.0054	0.7753 ± 0.0049	0.2528 ± 0.0016	0.6705 ± 0.0104	0.1679 ± 0.0027
(-) Dropout	0.6748 ± 0.0055	0.2549 ± 0.0025	0.3072 ± 0.0069	0.2601 ± 0.0046	–	0.7431 ± 0.0185	0.2405 ± 0.0047	0.7893 ± 0.0052	0.1641 ± 0.0043
(-) RC	–	–	0.2734 ± 0.0036	0.1948 ± 0.0096	0.8916 ± 0.0048	–	–	0.7520 ± 0.0157	0.1785 ± 0.0029
(-) FFN	–	–	0.2786 ± 0.0068	0.2314 ± 0.0073	0.9118 ± 0.0078	0.7432 ± 0.0052	0.2621 ± 0.0019	0.7672 ± 0.0071	0.1594 ± 0.0020
(-) PE	0.7069 ± 0.0093	0.2447 ± 0.0015	–	–	–	0.7593 ± 0.0051	0.2667 ± 0.0034	–	–
GIN$^+$	**0.7059** ± 0.0089	**0.2429** ± 0.0019	**0.3189** ± 0.0105	**0.2483** ± 0.0046	**0.9325** ± 0.0040	**0.7928** ± 0.0099	**0.2703** ± 0.0024	**0.8107** ± 0.0053	**0.1803** ± 0.0019
(-) Edge.	0.7033 ± 0.0015	0.2442 ± 0.0028	0.2956 ± 0.0047	0.2259 ± 0.0053	0.9286 ± 0.0049	0.7597 ± 0.0103	0.2702 ± 0.0021	0.2789 ± 0.0031	0.1752 ± 0.0020
(-) Norm	0.6934 ± 0.0077	0.2444 ± 0.0015	0.2707 ± 0.0037	0.2244 ± 0.0063	0.9322 ± 0.0025	0.7874 ± 0.0114	0.2556 ± 0.0026	0.6484 ± 0.0246	0.1722 ± 0.0034
(-) Dropout	0.6384 ± 0.0094	0.2531 ± 0.0030	0.3153 ± 0.0113	–	–	–	0.2545 ± 0.0068	0.7673 ± 0.0059	0.1730 ± 0.0018
(-) RC	0.6975 ± 0.0038	0.2527 ± 0.0015	0.2350 ± 0.0044	0.1741 ± 0.0085	0.9150 ± 0.0047	0.7733 ± 0.0122	0.1454 ± 0.0061	–	0.1617 ± 0.0026
(-) FFN	–	–	0.2393 ± 0.0049	0.1599 ± 0.0081	0.8944 ± 0.0074	–	0.2534 ± 0.0033	0.6676 ± 0.0039	0.1491 ± 0.0016
(-) PE	0.6855 ± 0.0027	0.2455 ± 0.0019	0.3141 ± 0.0031	–	–	0.7791 ± 0.0268	0.2601 ± 0.0023	–	–
GatedGCN$^+$	**0.7006** ± 0.0033	**0.2431** ± 0.0020	**0.4263** ± 0.0057	**0.3802** ± 0.0015	**0.9460** ± 0.0057	**0.8040** ± 0.0164	**0.2981** ± 0.0024	**0.8258** ± 0.0055	**0.1896** ± 0.0024
(-) Edge.	0.6882 ± 0.0028	0.2466 ± 0.0018	0.3764 ± 0.0117	0.3172 ± 0.0109	0.9372 ± 0.0062	0.7831 ± 0.0157	0.2951 ± 0.0028	0.0948 ± 0.0000	0.1891 ± 0.0021
(-) Norm	0.6733 ± 0.0026	0.2474 ± 0.0015	0.3628 ± 0.0043	0.3527 ± 0.0051	0.9326 ± 0.0056	0.7879 ± 0.0178	0.2748 ± 0.0012	0.6864 ± 0.0165	0.1743 ± 0.0026
(-) Dropout	0.6695 ± 0.0101	0.2508 ± 0.0014	0.3389 ± 0.0066	0.3393 ± 0.0051	–	–	0.2582 ± 0.0036	0.8088 ± 0.0062	0.1724 ± 0.0027
(-) RC	–	0.2498 ± 0.0034	0.4075 ± 0.0052	0.3475 ± 0.0064	0.9402 ± 0.0054	0.7833 ± 0.0177	0.2897 ± 0.0016	0.8099 ± 0.0053	0.1844 ± 0.0025
(-) FFN	–	–	–	0.3508 ± 0.0049	0.9364 ± 0.0059	–	0.2875 ± 0.0022	–	0.1718 ± 0.0024
(-) PE	0.6729 ± 0.0084	0.2461 ± 0.0025	0.4052 ± 0.0031	–	–	0.7771 ± 0.0057	0.2813 ± 0.0022	–	–

Ablation Studies - Dropout

- **Dropout proves advantageous for most datasets, with a very low dropout rate being sufficient and optimal.**

Table 6. Ablation study on LRGB and OGB datasets. - indicates that the corresponding hyperparameter is not used in GNN$^+$, as it empirically leads to inferior performance.

Metric	Peptides-func Avg. Precision ↑	Peptides-struct MAE ↓	PascalVOC-SP F1 score ↑	COCO-SP F1 score ↑	MalNet-Tiny Accuracy ↑	ogbg-molhiv AUROC ↑	ogbg-molpcba Avg. Precision ↑	ogbg-ppa Accuracy ↑	ogbg-code2 F1 score ↑
GCN$^+$	**0.7261** ± 0.0067	**0.2421** ± 0.0016	**0.3357** ± 0.0087	**0.2733** ± 0.0041	**0.9354** ± 0.0045	**0.8012** ± 0.0124	**0.2721** ± 0.0046	**0.8077** ± 0.0041	**0.1787** ± 0.0026
(-) Edge.	0.7191 ± 0.0036	–	0.2942 ± 0.0043	0.2219 ± 0.0060	0.9292 ± 0.0034	0.7714 ± 0.0204	0.2628 ± 0.0019	0.2994 ± 0.0062	0.1785 ± 0.0033
(-) Norm	0.7107 ± 0.0027	0.2509 ± 0.0026	0.1802 ± 0.0111	0.2332 ± 0.0079	0.9236 ± 0.0054	0.7753 ± 0.0049	0.2528 ± 0.0016	0.6705 ± 0.0104	0.1679 ± 0.0027
(-) Dropout	0.6748 ± 0.0055	0.2549 ± 0.0025	0.3072 ± 0.0069	0.2601 ± 0.0046	–	0.7431 ± 0.0185	0.2405 ± 0.0047	0.7893 ± 0.0052	0.1641 ± 0.0043
(-) RC	–	–	0.2734 ± 0.0036	0.1948 ± 0.0096	0.8916 ± 0.0048	–	–	0.7520 ± 0.0157	0.1785 ± 0.0029
(-) FFN	–	–	0.2786 ± 0.0068	0.2314 ± 0.0073	0.9118 ± 0.0078	0.7432 ± 0.0052	0.2621 ± 0.0019	0.7672 ± 0.0071	0.1594 ± 0.0020
(-) PE	0.7069 ± 0.0093	0.2447 ± 0.0015	–	–	–	0.7593 ± 0.0051	0.2667 ± 0.0034	–	–
GIN$^+$	**0.7059** ± 0.0089	**0.2429** ± 0.0019	**0.3189** ± 0.0105	**0.2483** ± 0.0046	**0.9325** ± 0.0040	**0.7928** ± 0.0099	**0.2703** ± 0.0024	**0.8107** ± 0.0053	**0.1803** ± 0.0019
(-) Edge.	0.7033 ± 0.0015	0.2442 ± 0.0028	0.2956 ± 0.0047	0.2259 ± 0.0053	0.9286 ± 0.0049	0.7597 ± 0.0103	0.2702 ± 0.0021	0.2789 ± 0.0031	0.1752 ± 0.0020
(-) Norm	0.6934 ± 0.0077	0.2444 ± 0.0015	0.2707 ± 0.0037	0.2244 ± 0.0063	0.9322 ± 0.0025	0.7874 ± 0.0114	0.2556 ± 0.0026	0.6484 ± 0.0246	0.1722 ± 0.0034
(-) Dropout	0.6384 ± 0.0094	0.2531 ± 0.0030	0.3153 ± 0.0113	–	–	–	0.2545 ± 0.0068	0.7673 ± 0.0059	0.1730 ± 0.0018
(-) RC	0.6975 ± 0.0038	0.2527 ± 0.0015	0.2350 ± 0.0044	0.1741 ± 0.0085	0.9150 ± 0.0047	0.7733 ± 0.0122	0.1454 ± 0.0061	–	0.1617 ± 0.0026
(-) FFN	–	–	0.2393 ± 0.0049	0.1599 ± 0.0081	0.8944 ± 0.0074	–	0.2534 ± 0.0033	0.6676 ± 0.0039	0.1491 ± 0.0016
(-) PE	0.6855 ± 0.0027	0.2455 ± 0.0019	0.3141 ± 0.0031	–	–	0.7791 ± 0.0268	0.2601 ± 0.0023	–	–
GatedGCN$^+$	**0.7006** ± 0.0033	**0.2431** ± 0.0020	**0.4263** ± 0.0057	**0.3802** ± 0.0015	**0.9460** ± 0.0057	**0.8040** ± 0.0164	**0.2981** ± 0.0024	**0.8258** ± 0.0055	**0.1896** ± 0.0024
(-) Edge.	0.6882 ± 0.0028	0.2466 ± 0.0018	0.3764 ± 0.0117	0.3172 ± 0.0109	0.9372 ± 0.0062	0.7831 ± 0.0157	0.2951 ± 0.0028	0.0948 ± 0.0000	0.1891 ± 0.0021
(-) Norm	0.6733 ± 0.0026	0.2474 ± 0.0015	0.3628 ± 0.0043	0.3527 ± 0.0051	0.9326 ± 0.0056	0.7879 ± 0.0178	0.2748 ± 0.0012	0.6864 ± 0.0165	0.1743 ± 0.0026
(-) Dropout	0.6695 ± 0.0101	0.2508 ± 0.0014	0.3389 ± 0.0066	0.3393 ± 0.0051	–	–	0.2582 ± 0.0036	0.8088 ± 0.0062	0.1724 ± 0.0027
(-) RC	–	0.2498 ± 0.0034	0.4075 ± 0.0052	0.3475 ± 0.0064	0.9402 ± 0.0054	0.7833 ± 0.0177	0.2897 ± 0.0016	0.8099 ± 0.0053	0.1844 ± 0.0025
(-) FFN	–	–	–	0.3508 ± 0.0049	0.9364 ± 0.0059	–	0.2875 ± 0.0022	–	0.1718 ± 0.0024
(-) PE	0.6729 ± 0.0084	0.2461 ± 0.0025	0.4052 ± 0.0031	–	–	0.7771 ± 0.0057	0.2813 ± 0.0022	–	–

Ablation Studies - Dropout

- **Dropout proves advantageous for most datasets, with a very low dropout rate being sufficient and optimal.**

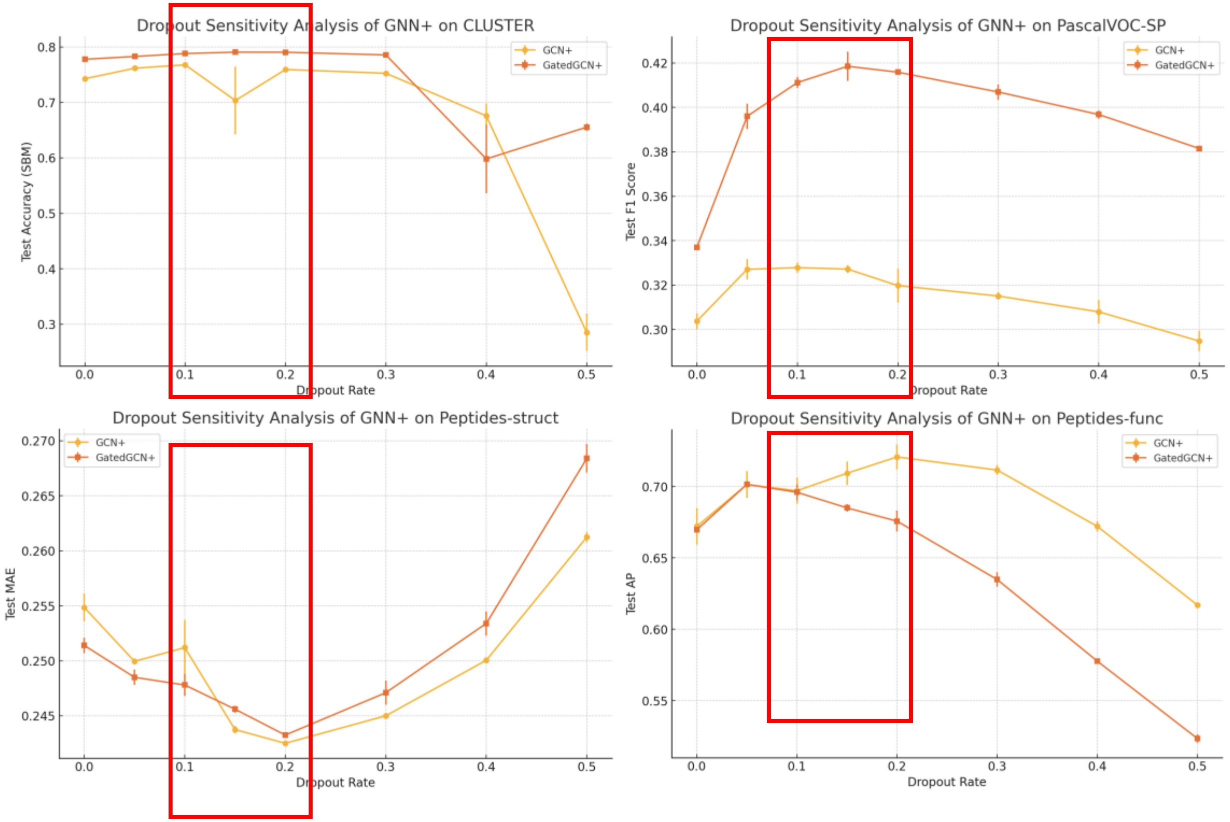


Figure 4. Sensitivity analysis of dropout rates in GNN^+ on CLUSTER, PascalVOC-SP, Peptides-struct, and Peptides-func.

Ablation Studies - Residual Connections

- **Residual connections are generally essential, except in shallow GNNs applied to small graphs.**

Table 6. Ablation study on LRGB and OGB datasets. - indicates that the corresponding hyperparameter is not used in GNN$^+$, as it empirically leads to inferior performance.

Metric	Peptides-func Avg. Precision ↑	Peptides-struct MAE ↓	PascalVOC-SP F1 score ↑	COCO-SP F1 score ↑	MalNet-Tiny Accuracy ↑	ogbg-molhiv AUROC ↑	ogbg-molpcba Avg. Precision ↑	ogbg-ppa Accuracy ↑	ogbg-code2 F1 score ↑
GCN$^+$	**0.7261** ± 0.0067	**0.2421** ± 0.0016	**0.3357** ± 0.0087	**0.2733** ± 0.0041	**0.9354** ± 0.0045	**0.8012** ± 0.0124	**0.2721** ± 0.0046	**0.8077** ± 0.0041	**0.1787** ± 0.0026
(-) Edge.	0.7191 ± 0.0036	–	0.2942 ± 0.0043	0.2219 ± 0.0060	0.9292 ± 0.0034	0.7714 ± 0.0204	0.2628 ± 0.0019	0.2994 ± 0.0062	0.1785 ± 0.0033
(-) Norm	0.7107 ± 0.0027	0.2509 ± 0.0026	0.1802 ± 0.0111	0.2332 ± 0.0079	0.9236 ± 0.0054	0.7753 ± 0.0049	0.2528 ± 0.0016	0.6705 ± 0.0104	0.1679 ± 0.0027
(-) Dropout	0.6748 ± 0.0055	0.2549 ± 0.0025	0.3072 ± 0.0069	0.2601 ± 0.0046	–	0.7431 ± 0.0185	0.2405 ± 0.0047	0.7893 ± 0.0052	0.1641 ± 0.0043
(-) RC	–	–	0.2734 ± 0.0036	0.1948 ± 0.0096	0.8916 ± 0.0048	–	–	0.7520 ± 0.0157	0.1785 ± 0.0029
(-) FFN	–	–	0.2786 ± 0.0068	0.2314 ± 0.0073	0.9118 ± 0.0078	0.7432 ± 0.0052	0.2621 ± 0.0019	0.7672 ± 0.0071	0.1594 ± 0.0020
(-) PE	0.7069 ± 0.0093	0.2447 ± 0.0015	–	–	–	0.7593 ± 0.0051	0.2667 ± 0.0034	–	–
GIN$^+$	**0.7059** ± 0.0089	**0.2429** ± 0.0019	**0.3189** ± 0.0105	**0.2483** ± 0.0046	**0.9325** ± 0.0040	**0.7928** ± 0.0099	**0.2703** ± 0.0024	**0.8107** ± 0.0053	**0.1803** ± 0.0019
(-) Edge.	0.7033 ± 0.0015	0.2442 ± 0.0028	0.2956 ± 0.0047	0.2259 ± 0.0053	0.9286 ± 0.0049	0.7597 ± 0.0103	0.2702 ± 0.0021	0.2789 ± 0.0031	0.1752 ± 0.0020
(-) Norm	0.6934 ± 0.0077	0.2444 ± 0.0015	0.2707 ± 0.0037	0.2244 ± 0.0063	0.9322 ± 0.0025	0.7874 ± 0.0114	0.2556 ± 0.0026	0.6484 ± 0.0246	0.1722 ± 0.0034
(-) Dropout	0.6384 ± 0.0094	0.2531 ± 0.0030	0.3153 ± 0.0113	–	–	–	0.2545 ± 0.0068	0.7673 ± 0.0059	0.1730 ± 0.0018
(-) RC	0.6975 ± 0.0038	0.2527 ± 0.0015	0.2350 ± 0.0044	0.1741 ± 0.0085	0.9150 ± 0.0047	0.7733 ± 0.0122	0.1454 ± 0.0061	–	0.1617 ± 0.0026
(-) FFN	–	–	0.2393 ± 0.0049	0.1599 ± 0.0081	0.8944 ± 0.0074	–	0.2534 ± 0.0033	0.6676 ± 0.0039	0.1491 ± 0.0016
(-) PE	0.6855 ± 0.0027	0.2455 ± 0.0019	0.3141 ± 0.0031	–	–	0.7791 ± 0.0268	0.2601 ± 0.0023	–	–
GatedGCN$^+$	**0.7006** ± 0.0033	**0.2431** ± 0.0020	**0.4263** ± 0.0057	**0.3802** ± 0.0015	**0.9460** ± 0.0057	**0.8040** ± 0.0164	**0.2981** ± 0.0024	**0.8258** ± 0.0055	**0.1896** ± 0.0024
(-) Edge.	0.6882 ± 0.0028	0.2466 ± 0.0018	0.3764 ± 0.0117	0.3172 ± 0.0109	0.9372 ± 0.0062	0.7831 ± 0.0157	0.2951 ± 0.0028	0.0948 ± 0.0000	0.1891 ± 0.0021
(-) Norm	0.6733 ± 0.0026	0.2474 ± 0.0015	0.3628 ± 0.0043	0.3527 ± 0.0051	0.9326 ± 0.0056	0.7879 ± 0.0178	0.2748 ± 0.0012	0.6864 ± 0.0165	0.1743 ± 0.0026
(-) Dropout	0.6695 ± 0.0101	0.2508 ± 0.0014	0.3389 ± 0.0066	0.3393 ± 0.0051	–	–	0.2582 ± 0.0036	0.8088 ± 0.0062	0.1724 ± 0.0027
(-) RC	–	0.2498 ± 0.0034	0.4075 ± 0.0052	0.3475 ± 0.0064	0.9402 ± 0.0054	0.7833 ± 0.0177	0.2897 ± 0.0016	0.8099 ± 0.0053	0.1844 ± 0.0025
(-) FFN	–	–	–	0.3508 ± 0.0049	0.9364 ± 0.0059	–	0.2875 ± 0.0022	–	0.1718 ± 0.0024
(-) PE	0.6729 ± 0.0084	0.2461 ± 0.0025	0.4052 ± 0.0031	–	–	0.7771 ± 0.0057	0.2813 ± 0.0022	–	–

Ablation Studies - FFN

- FFN is crucial for GIN+ and GCN+, greatly impacting their performance across datasets.

Table 6. Ablation study on LRGB and OGB datasets. - indicates that the corresponding hyperparameter is not used in GNN$^+$, as it empirically leads to inferior performance.

Metric	Peptides-func Avg. Precision ↑	Peptides-struct MAE ↓	PascalVOC-SP F1 score ↑	COCO-SP F1 score ↑	MalNet-Tiny Accuracy ↑	ogbg-molhiv AUROC ↑	ogbg-molpcba Avg. Precision ↑	ogbg-ppa Accuracy ↑	ogbg-code2 F1 score ↑
GCN$^+$	**0.7261** $\pm$ 0.0067	**0.2421** $\pm$ 0.0016	**0.3357** $\pm$ 0.0087	**0.2733** $\pm$ 0.0041	**0.9354** $\pm$ 0.0045	**0.8012** $\pm$ 0.0124	**0.2721** $\pm$ 0.0046	**0.8077** $\pm$ 0.0041	**0.1787** $\pm$ 0.0026
(-) Edge.	0.7191 $\pm$ 0.0036	–	0.2942 $\pm$ 0.0043	0.2219 $\pm$ 0.0060	0.9292 $\pm$ 0.0034	0.7714 $\pm$ 0.0204	0.2628 $\pm$ 0.0019	0.2994 $\pm$ 0.0062	0.1785 $\pm$ 0.0033
(-) Norm	0.7107 $\pm$ 0.0027	0.2509 $\pm$ 0.0026	0.1802 $\pm$ 0.0111	0.2332 $\pm$ 0.0079	0.9236 $\pm$ 0.0054	0.7753 $\pm$ 0.0049	0.2528 $\pm$ 0.0016	0.6705 $\pm$ 0.0104	0.1679 $\pm$ 0.0027
(-) Dropout	0.6748 $\pm$ 0.0055	0.2549 $\pm$ 0.0025	0.3072 $\pm$ 0.0069	0.2601 $\pm$ 0.0046	–	0.7431 $\pm$ 0.0185	0.2405 $\pm$ 0.0047	0.7893 $\pm$ 0.0052	0.1641 $\pm$ 0.0043
(-) RC	–	–	0.2734 $\pm$ 0.0036	0.1948 $\pm$ 0.0096	0.8916 $\pm$ 0.0048	–	–	0.7520 $\pm$ 0.0157	0.1785 $\pm$ 0.0029
(-) FFN	–	–	0.2786 $\pm$ 0.0068	0.2314 $\pm$ 0.0073	0.9118 $\pm$ 0.0078	0.7432 $\pm$ 0.0052	0.2621 $\pm$ 0.0019	0.7672 $\pm$ 0.0071	0.1594 $\pm$ 0.0020
(-) PE	0.7069 $\pm$ 0.0093	0.2447 $\pm$ 0.0015	–	–	–	0.7593 $\pm$ 0.0051	0.2667 $\pm$ 0.0034	–	–
GIN$^+$	**0.7059** $\pm$ 0.0089	**0.2429** $\pm$ 0.0019	**0.3189** $\pm$ 0.0105	**0.2483** $\pm$ 0.0046	**0.9325** $\pm$ 0.0040	**0.7928** $\pm$ 0.0099	**0.2703** $\pm$ 0.0024	**0.8107** $\pm$ 0.0053	**0.1803** $\pm$ 0.0019
(-) Edge.	0.7033 $\pm$ 0.0015	0.2442 $\pm$ 0.0028	0.2956 $\pm$ 0.0047	0.2259 $\pm$ 0.0053	0.9286 $\pm$ 0.0049	0.7597 $\pm$ 0.0103	0.2702 $\pm$ 0.0021	0.2789 $\pm$ 0.0031	0.1752 $\pm$ 0.0020
(-) Norm	0.6934 $\pm$ 0.0077	0.2444 $\pm$ 0.0015	0.2707 $\pm$ 0.0037	0.2244 $\pm$ 0.0063	0.9322 $\pm$ 0.0025	0.7874 $\pm$ 0.0114	0.2556 $\pm$ 0.0026	0.6484 $\pm$ 0.0246	0.1722 $\pm$ 0.0034
(-) Dropout	0.6384 $\pm$ 0.0094	0.2531 $\pm$ 0.0030	0.3153 $\pm$ 0.0113	–	–	–	0.2545 $\pm$ 0.0068	0.7673 $\pm$ 0.0059	0.1730 $\pm$ 0.0018
(-) RC	0.6975 $\pm$ 0.0038	0.2527 $\pm$ 0.0015	0.2350 $\pm$ 0.0044	0.1741 $\pm$ 0.0085	0.9150 $\pm$ 0.0047	0.7733 $\pm$ 0.0122	0.1454 $\pm$ 0.0061	–	0.1617 $\pm$ 0.0026
(-) FFN	–	–	0.2393 $\pm$ 0.0049	0.1599 $\pm$ 0.0081	0.8944 $\pm$ 0.0074	–	0.2534 $\pm$ 0.0033	0.6676 $\pm$ 0.0039	0.1491 $\pm$ 0.0016
(-) PE	0.6855 $\pm$ 0.0027	0.2455 $\pm$ 0.0019	0.3141 $\pm$ 0.0031	–	–	0.7791 $\pm$ 0.0268	0.2601 $\pm$ 0.0023	–	–
GatedGCN$^+$	**0.7006** $\pm$ 0.0033	**0.2431** $\pm$ 0.0020	**0.4263** $\pm$ 0.0057	**0.3802** $\pm$ 0.0015	**0.9460** $\pm$ 0.0057	**0.8040** $\pm$ 0.0164	**0.2981** $\pm$ 0.0024	**0.8258** $\pm$ 0.0055	**0.1896** $\pm$ 0.0024
(-) Edge.	0.6882 $\pm$ 0.0028	0.2466 $\pm$ 0.0018	0.3764 $\pm$ 0.0117	0.3172 $\pm$ 0.0109	0.9372 $\pm$ 0.0062	0.7831 $\pm$ 0.0157	0.2951 $\pm$ 0.0028	0.0948 $\pm$ 0.0000	0.1891 $\pm$ 0.0021
(-) Norm	0.6733 $\pm$ 0.0026	0.2474 $\pm$ 0.0015	0.3628 $\pm$ 0.0043	0.3527 $\pm$ 0.0051	0.9326 $\pm$ 0.0056	0.7879 $\pm$ 0.0178	0.2748 $\pm$ 0.0012	0.6864 $\pm$ 0.0165	0.1743 $\pm$ 0.0026
(-) Dropout	0.6695 $\pm$ 0.0101	0.2508 $\pm$ 0.0014	0.3389 $\pm$ 0.0066	0.3393 $\pm$ 0.0051	–	–	0.2582 $\pm$ 0.0036	0.8088 $\pm$ 0.0062	0.1724 $\pm$ 0.0027
(-) RC	–	0.2498 $\pm$ 0.0034	0.4075 $\pm$ 0.0052	0.3475 $\pm$ 0.0064	0.9402 $\pm$ 0.0054	0.7833 $\pm$ 0.0177	0.2897 $\pm$ 0.0016	0.8099 $\pm$ 0.0053	0.1844 $\pm$ 0.0025
(-) FFN	–	–	–	0.3508 $\pm$ 0.0049	0.9364 $\pm$ 0.0059	–	0.2875 $\pm$ 0.0022	–	0.1718 $\pm$ 0.0024
(-) PE	0.6729 $\pm$ 0.0084	0.2461 $\pm$ 0.0025	0.4052 $\pm$ 0.0031	–	–	0.7771 $\pm$ 0.0057	0.2813 $\pm$ 0.0022	–	–

Ablation Studies - PE

- PE is particularly effective for small-scale datasets, but negligible for large-scale datasets.

Table 5. Ablation study on GNN Benchmark (Dwivedi et al., 2023) (%). - indicates that the corresponding hyperparameter is not used in GNN$^+$, as it empirically leads to inferior performance.

Metric	ZINC MAE ↓	MNIST Accuracy ↑	CIFAR10 Accuracy ↑	PATTERN Accuracy ↑	CLUSTER Accuracy ↑
GCN$^+$	**0.076** $\pm$ 0.009	**98.382** $\pm$ 0.095	**69.824** $\pm$ 0.413	**87.021** $\pm$ 0.095	**77.109** $\pm$ 0.872
(-) Edge.	0.135 $\pm$ 0.004	98.153 $\pm$ 0.042	68.256 $\pm$ 0.357	86.854 $\pm$ 0.054	–
(-) Norm	0.107 $\pm$ 0.011	97.886 $\pm$ 0.066	60.765 $\pm$ 0.829	52.769 $\pm$ 0.874	16.563 $\pm$ 0.134
(-) Dropout	–	97.897 $\pm$ 0.071	65.693 $\pm$ 0.461	86.764 $\pm$ 0.045	74.926 $\pm$ 0.469
(-) RC	0.159 $\pm$ 0.016	95.929 $\pm$ 0.169	58.186 $\pm$ 0.295	86.059 $\pm$ 0.274	16.508 $\pm$ 0.615
(-) FFN	0.132 $\pm$ 0.021	97.174 $\pm$ 0.063	63.573 $\pm$ 0.346	86.746 $\pm$ 0.088	72.606 $\pm$ 1.243
(-) PE	0.127 $\pm$ 0.010	–	–	85.597 $\pm$ 0.241	75.568 $\pm$ 1.147
GIN$^+$	**0.065** $\pm$ 0.004	**98.285** $\pm$ 0.103	**69.592** $\pm$ 0.287	**86.842** $\pm$ 0.048	**74.794** $\pm$ 0.213
(-) Edge.	0.122 $\pm$ 0.009	97.655 $\pm$ 0.075	68.196 $\pm$ 0.107	86.714 $\pm$ 0.036	65.895 $\pm$ 3.425
(-) Norm	0.096 $\pm$ 0.006	97.695 $\pm$ 0.065	64.918 $\pm$ 0.059	86.815 $\pm$ 0.855	72.119 $\pm$ 0.359
(-) Dropout	–	98.214 $\pm$ 0.064	66.638 $\pm$ 0.873	86.836 $\pm$ 0.053	73.316 $\pm$ 0.355
(-) RC	0.137 $\pm$ 0.031	97.675 $\pm$ 0.175	64.910 $\pm$ 0.102	86.645 $\pm$ 0.125	16.800 $\pm$ 0.088
(-) FFN	0.104 $\pm$ 0.003	11.350 $\pm$ 0.008	60.582 $\pm$ 0.395	58.511 $\pm$ 0.016	62.175 $\pm$ 2.895
(-) PE	0.123 $\pm$ 0.014	–	–	86.592 $\pm$ 0.049	73.925 $\pm$ 0.165
GatedGCN$^+$	**0.077** $\pm$ 0.005	**98.712** $\pm$ 0.137	**77.218** $\pm$ 0.381	**87.029** $\pm$ 0.037	**79.128** $\pm$ 0.235
(-) Edge.	0.119 $\pm$ 0.001	98.085 $\pm$ 0.045	72.128 $\pm$ 0.275	86.879 $\pm$ 0.017	76.075 $\pm$ 0.845
(-) Norm	0.088 $\pm$ 0.003	98.275 $\pm$ 0.045	71.995 $\pm$ 0.445	86.942 $\pm$ 0.023	78.495 $\pm$ 0.155
(-) Dropout	0.089 $\pm$ 0.003	98.225 $\pm$ 0.095	70.383 $\pm$ 0.429	86.802 $\pm$ 0.034	77.597 $\pm$ 0.126
(-) RC	0.106 $\pm$ 0.002	98.442 $\pm$ 0.067	75.149 $\pm$ 0.155	86.845 $\pm$ 0.025	16.670 $\pm$ 0.307
(-) FFN	0.098 $\pm$ 0.005	98.438 $\pm$ 0.151	76.243 $\pm$ 0.131	86.935 $\pm$ 0.025	78.975 $\pm$ 0.145
(-) PE	0.174 $\pm$ 0.009	–	–	85.595 $\pm$ 0.065	77.515 $\pm$ 0.265

Table 6. Ablation study on LRGB and OGB datasets. - indicates that the corresponding hyperparameter is not used in GNN$^+$, as it empirically leads to inferior performance.

Metric	Peptides-func Avg. Precision ↑	Peptides-struct MAE ↓	PascalVOC-SP F1 score ↑	COCO-SP F1 score ↑	MalNet-Tiny Accuracy ↑	ogbg-molhiv AUROC ↑	ogbg-molpcba Avg. Precision ↑	ogbg-ppa Accuracy ↑	ogbg-code2 F1 score ↑
GCN$^+$	**0.7261** $\pm$ 0.0067	**0.2421** $\pm$ 0.0016	**0.3357** $\pm$ 0.0087	**0.2733** $\pm$ 0.0041	**0.9354** $\pm$ 0.0045	**0.8012** $\pm$ 0.0124	**0.2721** $\pm$ 0.0046	**0.8077** $\pm$ 0.0041	**0.1787** $\pm$ 0.0026
(-) Edge.	0.7191 $\pm$ 0.0036	–	0.2942 $\pm$ 0.0043	0.2219 $\pm$ 0.0060	0.9292 $\pm$ 0.0034	0.7714 $\pm$ 0.0204	0.2628 $\pm$ 0.0019	0.2994 $\pm$ 0.0062	0.1785 $\pm$ 0.0033
(-) Norm	0.7107 $\pm$ 0.0027	0.2509 $\pm$ 0.0026	0.1802 $\pm$ 0.0111	0.2332 $\pm$ 0.0079	0.9236 $\pm$ 0.0054	0.7753 $\pm$ 0.0049	0.2528 $\pm$ 0.0016	0.6705 $\pm$ 0.0104	0.1679 $\pm$ 0.0027
(-) Dropout	0.6748 $\pm$ 0.0055	0.2549 $\pm$ 0.0025	0.3072 $\pm$ 0.0069	0.2601 $\pm$ 0.0046	–	0.7431 $\pm$ 0.0185	0.2405 $\pm$ 0.0047	0.7893 $\pm$ 0.0052	0.1641 $\pm$ 0.0043
(-) RC	–	–	0.2734 $\pm$ 0.0036	0.1948 $\pm$ 0.0096	0.8916 $\pm$ 0.0048	–	–	0.7520 $\pm$ 0.0157	0.1785 $\pm$ 0.0029
(-) FFN	–	–	0.2786 $\pm$ 0.0068	0.2314 $\pm$ 0.0073	0.9118 $\pm$ 0.0078	0.7432 $\pm$ 0.0052	0.2621 $\pm$ 0.0019	0.7672 $\pm$ 0.0071	0.1594 $\pm$ 0.0020
(-) PE	0.7069 $\pm$ 0.0093	0.2447 $\pm$ 0.0015	–	–	–	0.7593 $\pm$ 0.0051	0.2667 $\pm$ 0.0034	–	–
GIN$^+$	**0.7059** $\pm$ 0.0089	**0.2429** $\pm$ 0.0019	**0.3189** $\pm$ 0.0105	**0.2483** $\pm$ 0.0046	**0.9325** $\pm$ 0.0040	**0.7928** $\pm$ 0.0099	**0.2703** $\pm$ 0.0024	**0.8107** $\pm$ 0.0053	**0.1803** $\pm$ 0.0019
(-) Edge.	0.7033 $\pm$ 0.0015	0.2442 $\pm$ 0.0028	0.2956 $\pm$ 0.0047	0.2259 $\pm$ 0.0053	0.9286 $\pm$ 0.0049	0.7597 $\pm$ 0.0103	0.2702 $\pm$ 0.0021	0.2789 $\pm$ 0.0031	0.1752 $\pm$ 0.0020
(-) Norm	0.6934 $\pm$ 0.0077	0.2444 $\pm$ 0.0015	0.2707 $\pm$ 0.0037	0.2244 $\pm$ 0.0063	0.9322 $\pm$ 0.0025	0.7874 $\pm$ 0.0114	0.2556 $\pm$ 0.0026	0.6484 $\pm$ 0.0246	0.1722 $\pm$ 0.0034
(-) Dropout	0.6384 $\pm$ 0.0094	0.2531 $\pm$ 0.0030	0.3153 $\pm$ 0.0113	–	–	–	0.2545 $\pm$ 0.0068	0.7673 $\pm$ 0.0059	0.1730 $\pm$ 0.0018
(-) RC	0.6975 $\pm$ 0.0038	0.2527 $\pm$ 0.0015	0.2350 $\pm$ 0.0044	0.1741 $\pm$ 0.0085	0.9150 $\pm$ 0.0047	0.7733 $\pm$ 0.0122	0.1454 $\pm$ 0.0061	–	0.1617 $\pm$ 0.0026
(-) FFN	–	–	0.2393 $\pm$ 0.0049	0.1599 $\pm$ 0.0081	0.8944 $\pm$ 0.0074	–	0.2534 $\pm$ 0.0033	0.6676 $\pm$ 0.0039	0.1491 $\pm$ 0.0016
(-) PE	0.6855 $\pm$ 0.0027	0.2455 $\pm$ 0.0019	0.3141 $\pm$ 0.0031	–	–	0.7791 $\pm$ 0.0268	0.2601 $\pm$ 0.0023	–	–
GatedGCN$^+$	**0.7006** $\pm$ 0.0033	**0.2431** $\pm$ 0.0020	**0.4263** $\pm$ 0.0057	**0.3802** $\pm$ 0.0015	**0.9460** $\pm$ 0.0057	**0.8040** $\pm$ 0.0164	**0.2981** $\pm$ 0.0024	**0.8258** $\pm$ 0.0055	**0.1896** $\pm$ 0.0024
(-) Edge.	0.6882 $\pm$ 0.0028	0.2466 $\pm$ 0.0018	0.3764 $\pm$ 0.0117	0.3172 $\pm$ 0.0109	0.9372 $\pm$ 0.0062	0.7831 $\pm$ 0.0157	0.2951 $\pm$ 0.0028	0.0948 $\pm$ 0.0000	0.1891 $\pm$ 0.0021
(-) Norm	0.6733 $\pm$ 0.0026	0.2474 $\pm$ 0.0015	0.3628 $\pm$ 0.0043	0.3527 $\pm$ 0.0051	0.9326 $\pm$ 0.0056	0.7879 $\pm$ 0.0178	0.2748 $\pm$ 0.0012	0.6864 $\pm$ 0.0165	0.1743 $\pm$ 0.0026
(-) Dropout	0.6695 $\pm$ 0.0101	0.2508 $\pm$ 0.0014	0.3389 $\pm$ 0.0066	0.3393 $\pm$ 0.0051	–	–	0.2582 $\pm$ 0.0036	0.8088 $\pm$ 0.0062	0.1724 $\pm$ 0.0027
(-) RC	–	0.2498 $\pm$ 0.0034	0.4075 $\pm$ 0.0052	0.3475 $\pm$ 0.0064	0.9402 $\pm$ 0.0054	0.7833 $\pm$ 0.0177	0.2897 $\pm$ 0.0016	0.8099 $\pm$ 0.0053	0.1844 $\pm$ 0.0025
(-) FFN	–	–	–	0.3508 $\pm$ 0.0049	0.9364 $\pm$ 0.0059	–	0.2875 $\pm$ 0.0022	–	0.1718 $\pm$ 0.0024
(-) PE	0.6729 $\pm$ 0.0084	0.2461 $\pm$ 0.0025	0.4052 $\pm$ 0.0031	–	–	0.7771 $\pm$ 0.0057	0.2813 $\pm$ 0.0022	–	–

Conclusions

- By integrating six widely used techniques into a unified GNN+ framework, we enhance three classic GNNs (GCN, GIN, and GatedGCN) for graph-level tasks.
- Evaluated on 14 datasets and fairly compared against 30 representative SOTA models proposed in the past three years, these classic GNNs rank Top-3 on all datasets and achieve the highest performance on 8 of them.

https://github.com/LUOyk1999/GNNPlus

Conclusions

- By integrating six widely used techniques into a unified GNN+ framework, we enhance three classic GNNs (GCN, GIN, and GatedGCN) for graph-level tasks.
- Evaluated on 14 datasets and fairly compared against 30 representative SOTA models proposed in the past three years, these classic GNNs rank Top-3 on all datasets and achieve the highest performance on 8 of them.

Thanks for listening!

https://github.com/LUOyk1999/GNNPlus