

## A Experimental Details

### A.1 Hyperparameters

For experiments of baselines and LC versions, we set hyperparameters similarly to [18],

- Activation: ReLU
- Batch size for precomputation-based GNNs: 4096
- Patience: 300
- Maximum number of epochs: 2000
- Hidden dimension: 256

We show hyperparameter search space for each GNN as follows.

**GCN.** The hyperparameter search space of GCN is listed as follows:

- Weight decay:  $[1e-8, 1e-7, 1e-6, 1e-5]$
- Learning rate:  $[0.005, 0.01, 0.02]$
- Dropout:  $[0, 0.5]$
- Number of layers:  $[2]$

**JKNet.** The hyperparameter search space of JKNet is listed as follows:

- Weight decay:  $[1e-8, 1e-7, 1e-6, 1e-5]$
- Learning rate:  $[0.001, 0.005, 0.01]$
- Dropout:  $[0, 0.5]$
- Number of layer:  $[3]$
- Pooling:  $[\text{concat}, \text{max}]$

**GPRGNN.** The hyperparameter search space of JKNet is listed as follows:

- Weight decay:  $[0, 1e-10, 1e-9, 1e-8]$
- Learning rate:  $[0.005, 0.01, 0.02]$
- Dropout for MLP:  $[0.5]$
- Dropout before propagation:  $[0, 0.5, 0.7]$
- Parameter initializing attention:  $[0, 0.1, 0.5, 0.9, 1]$
- Number of propagation layer:  $[5]$
- Number of layer of MLP:  $[4]$

**SGC.** The hyperparameter search space of SGC is listed as follows:

- Weight decay:  $[1e-9, 1e-8, 1e-7, 1e-6, 1e-5]$
- Learning rate:  $[1e-4, 0.001, 0.01]$
- Number of layers:  $[2]$

**FSGNN.** The hyperparameter search space of FSGNN is listed as follows:

- Weight decay1:  $[1e-5, 1e-4, 1e-3]$
- Learning rate1:  $[1e-4, 0.001, 0.01]$
- Weight decay2:  $[1e-6, 1e-5, 1e-4, 1e-3]$

- Learning rate2:  $[1e-4, 0.001, 0.01]$
- Weight decay3:  $[1e-6, 1e-5, 1e-4, 1e-3]$
- Learning rate3:  $[1e-4, 0.001, 0.01]$
- Weight decay for attention:  $[1e-6, 1e-5, 1e-4, 0.001]$
- Learning rate for attention:  $[0.001, 0.01, 0.1]$
- Dropout1:  $[0.5, 0.6, 0.7]$
- Dropout2:  $[0.5, 0.6, 0.7]$
- Number of layers:  $[3]$

**GCN<sub>LC</sub>.** The hyperparameter search space of GCN<sub>LC</sub> is listed as follows:

- Weight decay:  $[1e-7, 1e-6, 1e-5, 1e-4]$
- Learning rate:  $[5e-5, 1e-4, 0.001, 0.01]$
- Dropout:  $[0, 0.5]$
- Number of layers:  $[2]$

**JKNet<sub>LC</sub>.** The hyperparameter search space of JKNet<sub>LC</sub> is listed as follows:

- Weight decay:  $[1e-6, 1e-5, 1e-4, 0.001]$
- Learning rate:  $[1e-4, 0.001]$
- Dropout:  $[0, 0.5]$
- Number of layers:  $[3]$
- Pooling:  $[\text{concat}, \text{max}]$

**GPRGNN<sub>LC</sub>.** The hyperparameter search space of GPRGNN<sub>LC</sub> is listed as follows:

- Weight decay:  $[0, 1e-8, 1e-7]$
- Learning rate:  $[1e-4, 0.001, 0.01]$
- Dropout for MLP:  $[0.5]$
- Dropout before propagation:  $[0, 0.5, 0.7]$
- Parameter initializing attention:  $[0, 0.1, 0.5, 0.9, 1]$
- Number of propagation layer:  $[5]$
- Number of layer of MLP:  $[4]$

We choose the best parameter set from these candidates by utilizing Optuna [2] for 50 trials. Configuration of baselines and LC versions to reproduce experiments in our paper is shown in Table 5.

Table 5: Configuration for Table 2 and 4. OOM indicates that the execution is out of memory. “-” indicates that the parameter is not used for the GNN.

	Dataset	Weight decay	Learning rate	Dropout	Number of layers	Pooling	Attention initialing parameter							
GCN	Flickr	1e-6	0.02	0	2	-	-							
	Reddit	1e-7	0.02	0	2	-	-							
	arxiv	1e-5	0.02	0	2	-	-							
	papers100M	OOM	OOM	OOM	OOM	OOM	OOM							
JKNet	Flickr	1e-5	0.01	0	3	concat	-							
	Reddit	1e-7	0.005	0.5	3	concat	-							
	arxiv	1e-7	0.001	0.5	3	concat	-							
	papers100M	OOM	OOM	OOM	OOM	OOM	OOM							
GPRGNN	Flickr	0	0.005	MLP : 0.5 Propagation : 0.5	MLP: 4 Propagation : 5	-	0.1							
	Reddit	1e-9	1e-4	MLP : 0.5 Propagation : 0.5	MLP: 4 Propagation : 5	-	0.1							
	arxiv	1e-9	0.005	MLP : 0.5 Propagation : 0	MLP: 4 Propagation : 5	-	0.9							
	papers100M	OOM	OOM	OOM	OOM	OOM	OOM							
SGC	Flickr	1e-6	0.01	-	2	-	-							
	Reddit	1e-9	0.01	-	2	-	-							
	arxiv	1e-9	0.01	-	2	-	-							
	papers100M	1e-8	0.01	-	2	-	-							
FSGNN	Flickr	wd1 : 1e-4 wd2 : 1e-4 wd3 : 1e-4 wd.att : 0.01	lr1 : 1e-4 lr2 : 0.001 lr3 : 0.001 lr.att : 0.001	dp1 : 0.7 dp2 : 0.7	3	-	-							
		wd1 : 1e-5 wd2 : 0.001 wd3 : 1e-4 wd.att : 0.001	lr1 : 1e-4 lr2 : 0.001 lr3 : 0.001 lr.att : 0.01	dp1 : 0.6 dp2 : 0.5				3	-	-				
		wd1 : 1e-5 wd2 : 1e-5 wd3 : 1e-4 wd.att : 0.001	lr1 : 1e-4 lr2 : 0.001 lr3 : 0.001 lr.att : 0.01	dp1 : 0.7 dp2 : 0.6							3	-	-	
		wd1 : 1e-4 wd2 : 1e-5 wd3 : 1e-6 wd.att : 1e-4	lr1 : 1e-4 lr2 : 1e-4 lr3 : 0.001 lr.att : 0.1	dp1 : 0.5 dp2 : 0.5										3
	GCN.LC	Flickr	1e-5	1e-4	0	2	-							
	Reddit	1e-7	0.001	0.5	2	-	-							
	arxiv	1e-4	0.001	0	2	-	-							
	papers100M	1e-6	0.001	0	2	-	-							
JKNet.LC	Flickr	1e-4	1e-4	0	3	concat	-							
	Reddit	1e-6	1e-4	0.5	3	concat	-							
	arxiv	1e-6	1e-4	0.5	3	concat	-							
	papers100M	1e-6	0.001	0	3	concat	-							
GPRGNN.LC	Flickr	0	1e-4	MLP : 0.5 Propagation : 0	MLP: 4 Propagation : 5	-	0.1							
	Reddit	0	1e-4	MLP : 0.5 Propagation : 0	MLP: 4 Propagation : 5	-	0.5							
	arxiv	1e-8	0.001	MLP : 0.5 Propagation : 0.5	MLP: 4 Propagation : 5	-	0.9							
	papers100M	1e-7	0.001	MLP : 0.5 Propagation : 0	MLP: 4 Propagation : 5	-	1							