

## Model Configurations

In experiment 2, we validate the efficacy of our proposed MoDis by evaluating the performance of graph neural networks equipped with MoDis based pseudo-labeling. We employ grid search to determine the optimal hyper-parameters for both our method and base-lines. The specific hyper-parameter configurations can be found in Table 1, 2, 3, and 4. In these tables:

- "Hidden num" denotes the number of neurons in the network's hidden layer;
- "No. of layers" indicates the total number of layers within the network;
- "Learning rate" specifies the learning rate utilized;
- "Weight decay" denotes the weight decay associated with the L2 regularization term;
- "K" indicates the number of stage in the self-training algorithm;
- "P" denotes the total number of pseudo-labeled nodes used in each category;
- "Starting epoch" signifies the training epoch at which we start to record predictions to construct the memory bank;
- "Number of epochs" indicates the total number of training epochs during which predictions are recorded by the memory bank;

- " $\gamma$ " represent the temperature coefficient in the Sharpen function.

Experiment 3 aims to validate our proposed method on OOD (Out-of-Distribution) nodes in a graph. The hyper-parameters used in this experiment are summarized in Table 5. In this tables:

- "Hidden num" denotes the number of neurons in the network's hidden layer;
- "No. of layers" indicates the total number of layers within the network;
- "Learning rate" specifies the learning rate utilized;
- "Weight decay" denotes the weight decay associated with the L2 regularization term;
- "K" indicates the number of stage in the self-training algorithm;
- "P" denotes the total number of pseudo-labeled nodes used in each category;
- "Starting epoch" signifies the training epoch at which we start to record predictions to construct the memory bank;
- "Number of epochs" indicates the total number of training epochs during which predictions are recorded by the memory bank;
- " $\gamma$ " represent the temperature coefficient in the Sharpen function.

**Table 1: The hyper-parameters for GCN with the pseudo-labeling algorithm**

Dataset	Cora				Citeseer				Pubmed				CoraFull			
L/C	3	5	10	20	3	5	10	20	3	5	10	20	3	5	10	20
Hidden Num	16	16	16	16	16	16	16	16	16	16	16	16	64	64	64	64
No. of layers	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Learning Rate	5E-2	5E-2	5E-2	5E-2	5E-2	5E-2	5E-2	5E-2	5E-2	5E-2	5E-2	5E-2	5E-2	5E-2	5E-2	5E-2
Weight Decay	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4
K	2	2	2	2	7	7	3	5	5	5	2	3	3	3	3	3
Starting Epoch	0	20	50	50	50	50	50	20	50	50	50	50	100	100	100	100
Number of Epochs	150	100	100	100	150	100	100	50	100	100	100	100	100	100	100	100
P	174	192	112	112	223	214	90	61	560	700	677	480	30	12	15	6
$\gamma$	0.1	0.1	0.01	0.01	0.9	0.6	0.1	0.4	0.01	0.01	0.01	0.01	0.1	0.1	0.1	0.1

  

Dataset	AmazonCS				AmazonPhoto				CoauthorCS				CoauthorPhy			
L/C	3	5	10	20	3	5	10	20	3	5	10	20	3	5	10	20
Hidden Num	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64
No. of layers	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Learning Rate	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3
Weight Decay	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3
K	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Starting Epoch	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Number of Epochs	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
P	6	6	6	6	6	6	6	18	60	60	60	60	120	120	90	90
$\gamma$	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

**Table 2: The hyper-parameters for GAT with the pseudo-labeling algorithm**

Dataset	Cora				Citeseer				Pubmed				CoraFull			
L/C	3	5	10	20	3	5	10	20	3	5	10	20	3	5	10	20
Hidden Num	8	8	8	8	8	8	8	8	8	8	8	8	32	32	32	32
No. of layers	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Learning Rate	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-2	5E-2	5E-2	5E-2
Weight Decay	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4
K	3	4	4	3	6	3	2	2	4	4	3	2	3	3	3	3
Starting Epoch	20	100	10	50	50	50	50	50	50	50	50	20	100	100	100	100
Number of Epochs	100	150	150	100	100	100	100	100	100	100	100	100	100	100	100	100
P	96	150	48	42	212	160	84	60	480	600	320	36	30	12	15	6
$\gamma$	0.5	0.5	0.5	0.5	0.1	0.1	0.1	0.1	0.01	0.01	0.01	0.01	0.1	0.1	0.1	0.1

  

Dataset	AmazonCS				AmazonPhoto				CoauthorCS				CoauthorPhy			
L/C	3	5	10	20	3	5	10	20	3	5	10	20	3	5	10	20
Hidden Num	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32	32
No. of layers	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Learning Rate	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3
Weight Decay	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3
K	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Starting Epoch	100	100	100	150	150	150	150	100	100	100	100	100	100	100	100	100
Number of Epochs	100	100	100	50	50	50	50	100	100	100	100	100	100	100	100	100
P	6	6	6	3	3	12	3	12	60	100	100	100	100	100	100	100
$\gamma$	0.01	0.01	0.01	0.1	0.1	0.1	0.1	0.1	0.1	0.5	0.5	0.5	0.5	0.5	0.5	0.5

Table 3: The hyper-parameters for APPNP with the pseudo-labeling algorithm

Dataset	Cora				Citeseer				Pubmed				CoraFull			
L/C	3	5	10	20	3	5	10	20	3	5	10	20	3	5	10	20
Hidden Num	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64
No. of layers	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Learning Rate	1E-2	1E-2	1E-2	1E-2	1E-2	1E-2	1E-2	1E-2	1E-2	1E-2	1E-2	1E-2	5E-2	5E-2	5E-2	5E-2
Weight Decay	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4	5E-4
K	5	4	3	3	2	4	3	3	3	3	2	4	3	3	3	3
Starting Epoch	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Number of Epochs	100	100	100	200	50	50	50	50	100	100	50	100	100	100	100	100
P	111	178	63	63	111	96	96	54	675	720	708	328	30	12	15	6
$\gamma$	0.5	0.5	0.5	0.5	0.1	0.1	0.1	0.1	0.01	0.01	0.01	0.01	0.1	0.1	0.1	0.1

  

Dataset	AmazonCS				AmazonPhoto				CoauthorCS				CoauthorPhy			
L/C	3	5	10	20	3	5	10	20	3	5	10	20	3	5	10	20
Hidden Num	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64	64
No. of layers	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Learning Rate	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3
Weight Decay	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3	5E-3
K	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Starting Epoch	100	100	100	150	100	100	100	100	100	100	100	100	100	100	100	100
Number of Epochs	100	100	100	50	100	100	100	100	100	100	100	100	100	100	100	100
P	6	6	3	6	3	6	6	9	60	60	60	60	96	120	96	90
$\gamma$	0.1	0.1	0.01	0.01	0.4	0.4	0.4	0.4	0.01	0.1	0.1	0.1	0.01	0.1	0.1	0.1

**Table 4: The hyper-parameters for GCNII with the pseudo-labeling algorithm**

[illegible]

**Table 5: The hyper-parameters for OODGAT with the pseudo-labeling algorithm**

Dataset	Cora	AmazonCS	AmazonPhoto	CoauthorCS
ID Classes	[4,2,5,6]	[8,1,2,7,6]	[3,4,5,2,0]	[5,11,10,7,14,8,12,6]
Splits	[3,10,1000]	[3,10,5000]	[3,10,3000]	[3,10,8000]
Continuous	False	False	False	False
Weight Consistent	2	2	3	4
Weight Entropy	0.05	0.05	0.10	0.05
Weight Discrepancy	0.005	0.005	0.005	0.005
Margin	0.6	0.4	0.4	0.6
Heads	4	4	4	4
K	3	3	3	3
P	12	4	6	6
Starting Epoch	20	50	100	30
Number of Epochs	250	500	600	250
$\gamma$	0.3	0.1	0.01	0.3