1 Reproducibility Information

For all datasets, the data is split to a development and test set. Development set contains 1500 nodes, except Coauthor CS that contains 5000 nodes. Then, the development set is split to a training set containing 20 nodes for each class and a validation set with remaining nodes. For each experiment, we calculate the accuracy on 100 runs with different random split of training set, validation set and test set. Each run also has different random initializations. For each experiment, the accuracy is calculated by bootstrapping the accuracy results of 100 runs, with 1000 samples and 95% confidence-interval.

For training details, we stay the same with GDC [2]. We use early stopping strategy with patience of 100 epochs. The patience is reset when a new best accuracy is achieved on the validation set. We use the Adam optimizer [1] with learning rate of 0.01, for both normal parameters and t.

For ADC and GADC, we set the expansion step of Taylor expansion $K = 10$. For ADC, we select the initial value of t from $\{1, 3, 5\}$ based on the accuracy on the validation set. For GADC, we initialize $\theta_k = \frac{1}{K} = 0.1$. All the other hyperparameters of ADC and GADC are **same** as the original model. For CiteSeer, GCN's weight decay (10.0) is too large, which will cause the training process stopping very early. This is the only case we use different hyperparameters for ADC and GADC. For a full list of hyperparameters, see Sec.1.1.

We test all experiments on GeForce RTX 3090 with 24 GB memory.

1.1 Hyperparameters

Table 1: Hyperparameters for GCN

Model	Dataset name	t	k	ϵ	λ_{L_2}	Learning rate	Dropout	Hidden dimension	Hidden depth
No diffusion	CORA	-	-	-	0.06	0.01	0.5	64	1
	CITESEER				10.0				
	PUBMED				0.03				
	COAUTHOR CS				0.06				
	AMZ COMP				0.03				
	AMZ PHOTO				0.03				
GDC	CORA	5	-	0.0001	0.09	0.01	0.5	64	1
	CITESEER	4	-	0.0009	10.0				1
	PUBMED	3	-	0.0001	0.04				1
	COAUTHOR CS	1	64	-	0.08				1
	AMZ COMP	5	-	0.0010	0.07				1
	AMZ PHOTO	3	-	0.0001	0.08				2
ADC	CORA	5	-	-	0.06	0.01	0.5	64	1
	CITESEER	5			3.00				
	PUBMED	5			0.03				
	COAUTHOR CS	3			0.06				
	AMZ COMP	1			0.03				
	AMZ PHOTO	1			0.03				
GADC	CORA	-	-	-	0.06	0.01	0.5	64	1
	CITESEER				3.00				
	PUBMED				0.03				
	COAUTHOR CS				0.06				
	AMZ COMP				0.03				
	AMZ PHOTO				0.03				

Table 2: Hyperparameters for JKNet

Model	Dataset name	t	k	ϵ	λ_{L_2}	Learning rate	Dropout	Aggregation	Hidden dimension	Hidden depth
No diffusion	CORA	-	-	-	0.04	0.01	0.5	Concatenation	64	3
	CITESEER				1.00					4
	PUBMED				0.05					2
	COAUTHOR CS				0.02					2
	AMZ COMP				0.03					2
	AMZ PHOTO				0.03					2
GDC	CORA	5	-	0.0001	0.09	0.01	0.5	Concatenation	64	2
	CITESEER	4	-	0.0009	1.00					
	PUBMED	3	-	0.0001	0.09					
	COAUTHOR CS	1	64	-	0.03					
	AMZ COMP	5	-	0.0010	0.07					
	AMZ PHOTO	3	-	0.0005	0.07					
ADC	CORA	1	-	-	0.04	0.01	0.5	Concatenation	64	3
	CITESEER	3			1.00					4
	PUBMED	3			0.05					2
	COAUTHOR CS	1			0.02					2
	AMZ COMP	1			0.03					2
	AMZ PHOTO	1			0.03					2
GADC	CORA	-	-	-	0.04	0.01	0.5	Concatenation	64	3
	CITESEER				1.00					4
	PUBMED				0.05					2
	COAUTHOR CS				0.02					2
	AMZ COMP				0.03					2
	AMZ PHOTO				0.03					2

Table 3: Hyperparameters for ARMA

Model	Dataset name	t	k	ϵ	λ_{L_2}	Learning rate	Dropout	ARMA Layers	ARMA stacks	Hidden dimension	Hidden depth
No diffusion	CORA	-	-	-	0.04	0.01	0.5	1	3	16	1
	CITESEER				0.08				3		
	PUBMED				0.00				2		
	COAUTHOR CS				0.02				2		
	AMZ COMP				0.01				3		
	AMZ PHOTO				0.01				3		
GDC	CORA	5	64	-	0.08	0.01	0.5	1	2	16	1
	CITESEER	5	64	-	0.08				3		
	PUBMED	3	-	0.0001	0.00				2		
	COAUTHOR CS	1	64	-	0.01				3		
	AMZ COMP	5	64	-	0.04				3		
	AMZ PHOTO	3	64	-	0.04				2		
ADC	CORA	5	-	-	0.04	0.01	0.5	1	3	16	1
	CITESEER	3			0.08				3		
	PUBMED	1			0.00				2		
	COAUTHOR CS	1			0.02				2		
	AMZ COMP	1			0.01				3		
	AMZ PHOTO	1			0.01				3		
GADC	CORA	-	-	-	0.04	0.01	0.5	1	3	16	1
	CITESEER				0.08				3		
	PUBMED				0.00				2		
	COAUTHOR CS				0.02				2		
	AMZ COMP				0.01				3		
	AMZ PHOTO				0.01				3		

2 Experiment Results

Results in tabular form are presented as follow.

Table 4: Average accuracy (%) on datasets with bootstap-estimated 95% confidence levels.

Dataset	Model	No diffusion	GDC	ADC	GADC
CORA	GCN	82.21 ± 0.30%	83.06 ± 0.28%	84.55 ± 0.29%	84.83 ± 0.24%
	JKNet	81.52 ± 0.35%	82.89 ± 0.25%	83.54 ± 0.33%	84.17 ± 0.28%
	ARMA	81.64 ± 0.26%	83.62 ± 0.28%	84.37 ± 0.37%	84.49 ± 0.32%
CITESEER	GCN	74.08 ± 0.38%	74.56 ± 0.30%	74.49 ± 0.36%	74.94 ± 0.37%
	JKNet	71.69 ± 0.41%	73.15 ± 0.41%	73.02 ± 0.36%	73.57 ± 0.36%
	ARMA	73.06 ± 0.37%	74.27 ± 0.37%	73.83 ± 0.39%	74.09 ± 0.33%
PUBMED	GCN	78.18 ± 0.45%	78.99 ± 0.34%	82.35 ± 0.38%	81.92 ± 0.28%
	JKNet	78.09 ± 0.45%	78.86 ± 0.37%	83.00 ± 0.29%	82.43 ± 0.27%
	ARMA	76.95 ± 0.44%	78.34 ± 0.69%	80.91 ± 0.68%	79.67 ± 0.73%
COAUTHOR CS	GCN	91.77 ± 0.08%	92.87 ± 0.07%	93.14 ± 0.06%	92.93 ± 0.06%
	JKNet	91.72 ± 0.08%	92.59 ± 0.09%	92.46 ± 0.07%	92.53 ± 0.07%
	ARMA	91.71 ± 0.08%	92.52 ± 0.10%	92.35 ± 0.09%	91.05 ± 0.48%
AMZ COMP	GCN	84.63 ± 0.28%	86.18 ± 0.27%	86.25 ± 0.22%	85.16 ± 0.30%
	JKNet	83.97 ± 0.34%	85.48 ± 0.31%	84.79 ± 0.22%	83.39 ± 0.36%
	ARMA	85.12 ± 0.29%	85.65 ± 0.29%	85.41 ± 0.26%	84.09 ± 0.33%
AMZ PHOTO	GCN	91.97 ± 0.28%	89.34 ± 0.27%	93.00 ± 0.26%	92.55 ± 0.29%
	JKNet	91.52 ± 0.30%	91.60 ± 0.28%	92.77 ± 0.29%	92.41 ± 0.29%
	ARMA	91.89 ± 0.29%	92.34 ± 0.27%	93.19 ± 0.25%	92.26 ± 0.28%

References

[1] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[2] Johannes Klicpera, Stefan Weiß enberger, and Stephan Günnemann. Diffusion improves graph learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.