AUTOCOG: A UNIFIED DATA-MODAL CO-SEARCH FRAMEWORK FOR GRAPH NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

A APPENDIX

In this section, we outline the derived architectures from **AutoCoG** at all configurations for all datasets. Below you can find our experimental settings; next to it are the three tables, each outlines the architecture configurations for two layers, 16 layers, and 32 layers.

Experimental settings. We use the Adam optimizer for both model and architecture weights with a learning rate of 0.01, and 0.001 respectively. L_2 regularization for the model's weight is set to be 0.0005 as standard for Cora/Citeseer/Pubmed, and zero for ogbn-arxiv. We set dropout rates to be 0.6 for Cora and Citeseer; 0.5 for Pubmed; and 0.1 for ogbn-arxiv. For Identity-Mapping, γ is chosen to be 0.5 for Cora/Pubmed/ogbn-arxiv, and 0.6 for Citeseer. Following best practices, we set the hidden dimension to be 256 for Citeseer/Arvix, and 64 for Pubmed/Cora. Number of head is one for all. In addition, we set the η, ratio between initial connection and current features, to be 0.1. For P-DARTS, we set N to be seven with an increment of two layers per stage. Finally, we allow our search and train to reach a maximum of 1000 epochs, while setting our patience to be 400, and 200 respectively.

Table 1: architectures derived for our 2 layers settings. the left most setting are for layer 1, while the right most settings are for layer 2.

dataset	data policy (p)	attention	activation	skip
Cora	DropNode (0.15)	const, gat	tanh, tanh	None, None
Citeseer	DropNode (0.10)	cos, gcn	relu6, leaky relu	None, None
Pubmed	Identity	gcn, gcn	tanh, tanh	initial skip, initial skip
ogbn-arxiv	DropNode (0.15)	gcn, gcn	tanh, tanh	inital skip, inital skip

Table 2: architectures derived for our 16 layers settings. The same setting is repeated across the depth of our model

dataset	data policy (p)	attention	activation	skip
Cora	DropNode (0.2)	gcn	linear	initial skip
Citeseer	Identity	gcn	tanh	initial skip
Pubmed	AddEdge (0.1)	gcn	tanh	initial skip
ogbn-arxiv	DropNode (0.15)	gcn	tanh	inital skip

Table 3: architectures derived for our 32 layers settings. The same setting is repeated across the depth of our model

dataset	data policy (p)	attention	activation	skip
Cora	AddEdge (0.15)	gcn	elu	initial skip
Citeseer	AddEdge (0.10)	gcn	tanh	initial skip
Pubmed	AddEdge (0.10)	gcn	tanh	initial skip
ogbn-arxiv	DropNode (0.10)	gcn	tanh	inital skip