Experiment Details

No of layers	2		
First layer Attn heads	8	2nd layer attn heads	1
first layer feature/head	8	2nd layer feature o/p	6(classes)
Total features	8*8 = 64		
First layer activation funcn	elu	2nd layer activation func	softmax
Regularization	0.0005	learning rate	0.01
Dropout	0.6	epochs	200
Weight initialization	Xavier	early stopping	(alike GCN) 10

Optimization

	dropout	Optimizer	Epochs	Test Acc	Weight decay	Learning rate	first layer activation function
	0.6	Adam	200	68.2	0.0005	0.005	elu
	0.6	Adam	200	67.8	0.0005	0.01	elu
With Normalize data + early stopping	0.6	Adam	200	70.3	0.0005	0.01	elu
Xavier Weight initilazation	0.6	Adam	200	71.5	0.0005	0.01	elu
			39/200	71.8			relu
				71.5			
Seed = 42				72.1			
				Highest 72.60			