

Experiment Details									
Number of Layers	2	Learning Rate	0.01	Dropout rate	0.5				
Hidden units	16	epoch	200	Feature Normalization	Row-normalized features				
Output units	6	early stopping window	10						
Optimizer	Adam	L2 regularization factor	0.0005						
Loss function: F_HL_loss									
Optimization									
seed	Layers	hidden dim	dropout	Optimizer	Epochs	Test Acc	Weight decay	Learning rate	
	2	16	-	Adam	200	67.2	5.00E-04	0.01	
	2	16	0.5	Adam	200	68	5.00E-04	0.01	
	2	16	0.5	Adam	200	60	0		
	42	2	16	0.5	Adam	200	64	0	0.01
	42	2	16	0.5	Adam	200	65.5 (Best)	0	0.01
42	2	16	0.5	Adam	200	68.8	5.00E-04		
42	2	16	0.5	AdamW	200	66.1	5.00E-04		
42	2	16	0.4	Adam	200	66.9	5.00E-04		
42	2	16	0.5	Adam	200	67.9	5.00E-04		
				Adam	200	66.5	5.00E-04	0.005	
						66.6		0.05	
0						68.8		0.01	
data x = data x / data.x.sum(dim=1, keepdim=True).clamp(min=1e-10)									
With row normalization	2	16	0.5	Adam	200	75.4	5.00E-04	0.01	