Table 1: Ablation studies on two homophilic and two heterophilic datasets. Metrics: accuracy for Photo and Computer, ROC-AUC ($\times 100$) for Tolokers and Minesweeper. Spexphormer-uniform instead of using the attention scores, samples uniformly from node neighborhoods. Spexphormer w.o. temp and Spexphormer w.o. layer norm remove temperature and layer normalization from the attention score estimator network respectively. For the initial network, we report the result for the network used for training the Spexphormer, thus there is no confidence interval for them. We can see that uniform sampling consistently performs worse than using attention weights. Temperature and layer normalization generally improve results.

Model/Dataset	Computer	Photo	Minesweeper	Tolokers
Spexphormer	91.09 ± 0.08	95.24 ± 0.12	90.72 ± 0.06	83.34 ± 0.13
Spexphormer-uniform	86.65 ± 0.46	94.21 ± 0.22	84.15 ± 0.22	82.56 ± 0.17
Spexphormer w.o. temp	89.05 ± 0.35	95.30 ± 0.16	90.02 ± 0.02	83.34 ± 0.13
Spexphormer w.o. layer norm	89.70 ± 0.25	94.91 ± 0.18	89.65 ± 0.10	84.06 ± 0.10
Initial Network	85.23	91.70	85.67	80.16



Model	ogbn-proteins	Amazon 2M	Pokec		
MLP	72.04 ± 0.48	63.46 ± 0.10	60.15 ± 0.03		
GCN	72.51 ± 0.35	83.90 ± 0.10	62.31 ± 1.13		
SGC	70.31 ± 0.23	81.21 ± 0.12	52.03 ± 0.84		
GCN-NSAMPLER	73.51 ± 1.31	83.84 ± 0.42	63.75 ± 0.77		
GAT-NSAMPLER	74.63 ± 1.24	85.17 ± 0.32	62.32 ± 0.65		
SIGN	71.24 ± 0.46	80.98 ± 0.31	68.01 ± 0.25		
NodeFormer	77.45 ± 1.15	87.85 ± 0.24	70.32 ± 0.45		
SGFormer	79.53 ± 0.38	89.09 ± 0.10	73.76 ± 0.24		
Spexphormer	$\textbf{80.65} \pm \textbf{0.07}$	$\textbf{90.32} \pm \textbf{0.01}$	$\textbf{74.73} \pm \textbf{0.04}$		
Memory Information for Spexphormer					
Memory (MB)	2232	3094	2128		
Batch Size	256	500	500		
Hidden Dimension	128	128	64		
Number of layers	2	2	2		
Number of Parameters	297,592	300,209	83,781		

Figure 1: Memory usage comparison: Attention Score Estimator network and Spexphormer vs. enhanced Exphormer (with per-layer edge embeddings) for degrees 6 and 30. Exphormer with a degree of 30 could not fit into the memory of a 40GB GPU device, and thus the number is a lower bound here. Memory usage gap grows with graph size.

Table 2: Comparative results on large graph datasets, with $ROC-AUC(\times 100)$ reported for the ogbn-proteins dataset and accuracy for all others. GPU memory usage, batch sizes, hidden dimensions used to obtain these numbers, and the total number of parameters have been added at the bottom of the table.



Figure 2: The memory, run-time trade-off for ogbn-proteins and ogbn-arxiv datasets. It is worth mentioning that the experiments with different batch sizes yield similar results for test accuracy/AUC. Memory and time can be traded in our approach.