TABLE VIII
STATISTICS OF DATASETS.

Dataset	Task	Number of graph	Node type	Number of Nodes	Edge type	Number of edge	Meta-path
ACM	Node classification	1	Paper Author Subject	2094 4070 73	Paper-Paper Paper-Subject Paper-Author	4119 2094 6781	P-P P-S-P P-A-P
Movies	Node classification	1	Movie Director Starring	4283 1566 6806	Movie-Director Movie-Starring Director-Director Starring-Starring	3834 15693 4314 50870	M-D-M M-S-M
OGB-molbace	Graph classification	1513	C N O F S Cl Br I	39055 5453 4586 1721 499 218 41 4	C-C C-N C-O C-F C-S C-Cl C-Br C-I N-N N-O N-S O-O O-S F-F S-S Cl-Cl Br-Br I-I	111493 10707 5020 1721 722 218 41 4 5605 64 269 4586 707 1721 499 218 41 4	C-C C-O-C C-N-C
OGB-molbace	Graph regression	1128	C N O F S Cl Br I P	11207 970 1768 100 166 645 72 18 45	C-C C-O C-N C-Cl C-S N-O P-O C-F N-N C-Br S-O S-P S-N C-I S-S P-N C-P	20398 1897 1823 645 181 172 118 100 82 72 66 56 29 18 10 4 2	C-C C-O-C
ADNI	Graph classification	230	L R	28290 28290	Inter-hemispheric edges Intra-hemispheric edges	597212 851326	L-R-L R-L-R
PROTEINS	Graph classification	1025	Helice Sheet	20228 19525	Helice-Helice Sheet-Sheet	53430 47080	Helice-Sheet-Helice Sheet-Helice-Sheet

APPENDIX

A. Implementation details for each datasets

ACM: Since Author and Subject nodes do not have input features in this dataset, trainable embeddings with 128 dimensions are used as input node features. Features of Paper node are first projected to 128 dimensions aligned with other nodes via a linear layer. Hidden dimensions of all nodes are 128 and learning rate is 1e-3. L2 regularization is 1e-4, and head number is 1.

Movies: Different types of nodes have different dimensions of initial features. Therefore, they are all projected to 128 dimensions first. We obtain global representation for Hidden dimensions of all nodes are 64. Learning rate is 1e-3. L2 regularization is 1e-4, and head number is 1

OGB-molbace: Since Carbon atom is the key atom of whole graphs, we optimize our model with this atom. Besides, there are relatively larger numbers of Nitrogen and Oxygen atoms. Therefore, we take three meta-paths (C-C, C-O-C, C-N-C) into consideration in this work. Initial embedding dimensions are 100. Hidden dimensions of all nodes are 256 and the head number is 8 (i.e., dimension of single head is 32). Learning rate is 1e-3. Dropout is 0.3 and L2 regularization is 1e-4.

OGB-molesol: We take two meta-paths (C-C, C-O-C) into consideration in this work. Initial embedding dimensions are 32. Hidden dimensions of all nodes are 128 and the head number is 16 (i.e., dimension of single head is 8). Learning rate is 1e-3 and L2 regularization is 1e-4.

ADNI: We take two meta-paths (L-R-L and R-L-R) into consideration in this work. The brain networks are dense graphs. Therefore, we binary the network with a 90% quantile and the functional connectivity are modeled as initial node features for each subject. Initial embedding dimensions are 246 (the number of brain nodes). Hidden dimensions of all

nodes are 64 and the head number is 8. Learning rate is 1e-3 and L2 regularization is 1e-4.

PROTEINS: We take two meta-paths (sheet-helice-sheet and helice-sheet-helice) into consideration in this work. Initial embedding dimensions are 29. Hidden dimensions of all nodes are 128 and the head number is 8. Learning rate is 1e-3 and L2 regularization is 1e-5.