1 Data Statistics

The 11 graph datasets used in our study are collected from [2], including seven biological datasets and four social datasets. Their data statistics and properties are summarized in Table 1 and Table 2. Among the biological datasets, ENZYMES contains six classes, while the other datasets have only two classes.

Table 1: Data statistics of biological graphs.

Dataset	# Graphs	# Classes	Avg. # Nodes	Avg. # Edges	Node Labels	Node Attr (Dim)
MUTAG	188	2	17.93	19.79	+	–
NCI1	4110	2	29.87	32.30	+	–
PROTEINS	1113	2	39.06	72.82	+	+ (1)
D&D	1178	2	284.32	715.66	+	–
ENZYMES	600	6	32.63	62.14	+	+ (18)
PTC	344	2	14.29	14.69	+	–
NCI109	4127	2	29.68	32.13	+	–

Table 2: Data statistics of social graphs.

Dataset	# Graphs	# Classes	Avg. # Nodes	Avg. # Edges	Node Labels	Node Attr (Dim)
COLLAB	5000	3	74.49	2457.78	–	–
IMDB-B	1000	2	19.77	96.53	–	–
IMDB-M	1500	3	13.00	65.94	–	–
RE-B	2000	2	429.63	497.75	–	–

2 More Details on Experimental Settings

2.1 Node Representations

We follow the same setting in GFN [1] to establish the basic node representations by considering node degree and multi-scale propagated features. The obtained basic representations are then used for building disentangled graph capsules as stated in the main paper.

2.2 Hyper-parameters

Our model is trained using Adam optimizer with hyper-parameters detailed in Table 3 and Table 4, where d_1 is the feature dimension of primary graph capsules $\mathbf{u}^{(1)}$, N_2 is the number of capsules in the second capsule layer. Note, HGCN contains three capsule layers, the first layer is the primary graph capsule layer, the last layer is the class capsule layer. All of our experiments are implemented with PyTorch on a single GeForce GTX 1080 GPU which has 12GB memory.

Table 3: Hyper-parameters of biological datasets.

Dataset	Batch Size	d_1	N_2	Dropout Rate	Weight Decay	Learning Rate
MUTAG	16	128	5	0.7	1e-3	1e-3
NCI1	128	128	10	0.2	1e-4	1e-3
PROTEINS	128	128	25	0.7	1e-3	1e-3
D&D	64	32	20	0.7	1e-4	1e-3
ENZYMES	16	128	10	0.3	1e-4	1e-3
PTC	16	64	15	0.5	1e-3	1e-3
NCI109	512	128	16	0.1	1e-4	1e-3

Table 4: Hyper-parameters of social datasets.

Dataset	Batch Size	d_1	N_2	Dropout Rate	Weight Decay	Learning Rate
COLLAB	256	128	20	0.1	1e-4	1e-3
IMDB-B	256	128	10	0.1	1e-4	3e-3
IMDB-M	64	128	10	0	1e-4	3e-3
RE-B	16	128	15	0.1	1e-4	1e-3

3 Performance Comparison with C*

Table 5 and 6 show the performance comparison on biological and social graphs using the criterion C^*, respectively.

Table 5: Performance comparison on biological graphs (C^*).

Algorithm		**PROTEINS**	**D&D**	**ENZYMES**
STRUCTPOOL	GNN	80.36	84.19	63.83
MemGNN	GNN	81.35	82.92	75.50
GMN	GNN	**82.25**	**84.40**	78.66
Ours	GNN	81.32±2.23	84.04±2.38	**84.17±4.03**

Table 6: Performance comparison on social graphs (C^*).

Algorithm		**COLLAB**	**IMDB-B**	**IMDB-M**	**RE-B**
STRUCTPOOL	GNN	74.22	74.70	52.47	—
MemGNN	GNN	77.0	—	—	85.55
GMN	GNN	80.18	—	—	95.28
Ours	GNN	**84.80±1.57**	**79.80±3.39**	**55.80±2.20**	**95.30±1.30**

References

[1] T. Chen, S. Bian, and Y. Sun. Are powerful graph neural nets necessary? a dissection on graph classification. *International Conference on Learning Representations (ICLR Workshop)*, 2019.

[2] K. Kersting, N. M. Kriege, C. Morris, P. Mutzel, and M. Neumann. Benchmark data sets for graph kernels, 2016. URL `http://graphkernels.cs.tu-dortmund.de`.