Supporting Information for
Hyperparameter-free and Explainable Whole Graph Embedding

A Schematics

A.1 Schematics of computing degree, H-index and coreness

As shown in Figure. 1.

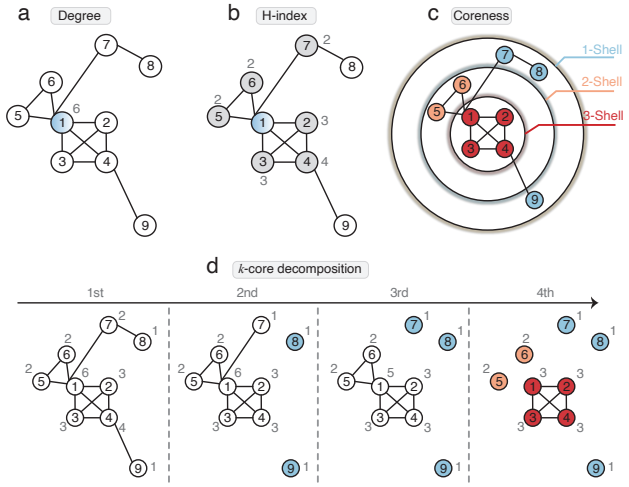

Figure 1: **Illustration of computing the nodal degree, H-index and coreness in a graph**. **a**, The degree of node 1 is six because it is connected with six nodes in the graph. **b**, After obtaining the degrees of the neighbors of node 1, it can be calculated that the H-index of node 1 is three. The same goes for other nodes in the graph. **c**, The coreness of nodes can be calculated by implementing the k-core decomposition on the graph. During the k-core decomposition process, each node is assigned a shell layer, called the node's coreness. The k-core decomposition process is illustrated in **d**. Specifically, the nodes with a degree of 1 (i.e., nodes 8 and 9) are initially removed. Since it reduces the degree of node 7 from $k = 2$ to $k = 1$, node 7 is also removed, until there are no remaining nodes with a degree of 1. All the removed nodes (i.e., nodes 7, 8, and 9) and their linked edges form the 1-shell. The same operation will be repeated until all nodes in the graph have been assigned a shell.

A.2 Schematic of DHC updating procedure

As shown in Figure. 2.

A.3 Schematic of computing Shannon entropy

Take the initial state in Fig. 2 as an example. Denote the H-index of a node as a random variable X. The probability

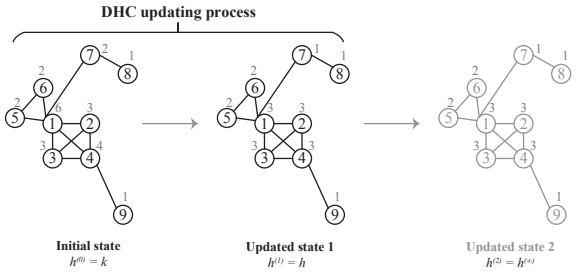

Figure 2: **Schematics of DHC updating procedure.** The zero-order H-index of every node is initialized to its degree, as shown in the initial state. The following updated states can be calculated by the DHC theorem. When it comes to the updated state 2, the second-order H-index of each node is the same as that of the previous step (i.e., the first-order, $h^{(1)}$). So, the iterative update has already converged and the convergence values are the nodal coreness, as shown in Fig. 1d.

distribution of H-indices in the initial state is $\mathrm{P}(X = 6) = \frac{1}{9}, \mathrm{P}(X = 3) = \frac{2}{9}, \mathrm{P}(X = 4) = \frac{1}{9}, \mathrm{P}(X = 2) = \frac{1}{3}, \mathrm{P}(X = 1) = \frac{2}{9}$. Based on it, one can measure the Shannon entropy of the initial state as $H_0 = \frac{1}{9}\log_2\frac{1}{9} + \frac{2}{9}\log_2\frac{2}{9} + \frac{1}{9}\log_2\frac{1}{9} + \frac{1}{3}\log_2\frac{1}{3} + \frac{2}{9}\log_2\frac{2}{9} \approx 2.197$. The same is true for other states of the graph in different iterations.

A.4 Schematics of DHC-E

As shown in Figure. 3.

B Experimental Datasets and Hyperparameters of Experimental Baselines

Simulation datasets were constructed by generating a certain number of specific networks, including *random network (ER)* [Erdős and Rényi, 1960], *small-world network (SW)* [Albert *et al.*, 1999], and *scale-free network (BA)* [Faloutsos *et al.*, 1999]). Each of them has unique structural characteristics and could be generated by specific algorithms, respectively. **Real-world datasets** were picked from online public resources-TUDatasets [Morris *et al.*, 2020] (https://chrsmrrs.github.io/datasets/docs/datasets/), used for graph classification. The

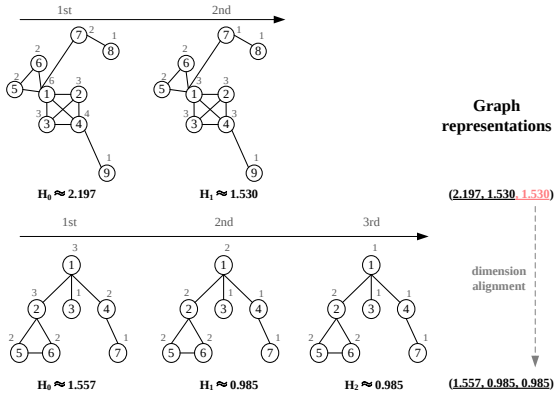

Figure 3: **Schematics of DHC-E.** For the other states generated from the DHC updating process of a graph, the same rules are applied to calculate their Shannon entropy. At the end of the DHC updating process, one obtains two Shannon entropy sequences of the two graphs $(2.197, 1.530)$ and $(1.557, 0.985, 0.985)$, which are two graphs' embeddings. The two embeddings are with unequal dimensions. Based on the highest dimension of the two embeddings (i.e., 3), the last element of $(2.197, 1.530)$ is used to expand $(2.197, 1.530)$ to $(2.197, 1.530, 1.530)$ for dimension alignment.

simulation datasets were more convenient for use to explore the performance of DHC-E based on datasets with different complexity because one could control their node-scale and network's sparsity. Although this way leads to different structures of ER, SW, and BA, their characteristics remain to be distinguishable from the other two categories. The real-world datasets directly evaluated the practical application value of DHC-E. This combination of simulated and real-world data allow for a more comprehensive evaluation of different models.

Besides, to evaluate the performance of the algorithm on intelligent biological systems, **Brain morphological similarity networks** were generated for autism spectrum disorders (ASD) and typical control (TC) participants, using the second iteration of Autism Brain Imaging Data Exchange (ABIDE II, http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html) [Di Martino *et al.*, 2017]. These data have been openly released to the scientific community on June 2016. In accordance with HIPAA guidelines and 1000 Functional Connectomes Project / INDI protocols, all datasets are anonymous, with no protected health information included. The current used data were downloaded from the Indiana University site (ABIDE II). Forty participants were enrolled with 20 ASD and 20 TC participants. The ASD participants' age range is 17-54 years and TC's age range is 19-37 years. Computational Anatomy Toolbox (CAT 12, http://www.neuro.uni-jena.de/cat/) for SPM (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) was used to perform the surface analysis to calculate the cortical sulcal depth. Then, the data were resampled to a lower 32k mesh (average node spacing of $\sim$2 mm) compatible with the Human Connectome Project (HCP) with a smoothing filter size of 25 mm. Specifically, to construct the brain network, nodes were first defined based on the Schaefer atlas [Schaefer *et al.*, 2018] with 400 brain regions ($Schaefer2018_400Parcels_17Networks$). Second, the relationships among different nodes were calculated by the Earth Mover's Distance (EMD) (D_E) between the probability distribution function (PDF) of the brain cortical sulcal depth of any paired brain regions [Li *et al.*, 2021; Wu *et al.*, 2018]. In detail, the brain cortical sulcal depth value was extracted for each brain region. Then, the kernel density estimation (KDE) was used to estimate the probability density function of these sulcal depth values with 128 sampling points [Wang *et al.*, 2016] for each brain region. The probability density function is standardized by dividing its sum to produce PDF for each brain region, and the EMD between the PDFs of any paired brain regions was used for constructing an EMD matrix (400×400). Subsequently, the EMD matrix was converted to a similarity matrix based on sigmoid function and termed the sulcal depth-based morphological similarity brain network M_{sim}, for each participant, as $M_{sim} = \sqrt{\frac{1}{1+D_E}}$. Finally, the M_{sim} was thresholded into a binary network by keeping the top 10% edges.

B.1 Datasets for binary classification

datasets	Sample size	Node-scale configuration	Sparsity configuration	Total size
B1	10	[100]	[0.1]	2×10×1=20
B2	10	[100]	[0.1,0.9]	2×10×2=40
B3	10	[100,200]	[0.1,0.9]	2×10×4=80

Table 1: **Description of simulation datasets for binary classification.**For example, the node-scale configuration and sparsity configuration of B3 is $[100, 200]$ and $[0.1, 0.9]$, which indicates that it has four possible configurations (i.e., $[100, 0.1]$, $[100, 0.9]$, $[200, 0.1]$, $[200, 0.9]$) for each category of the generated networks. Since there are ten samples for each category, the total sample size will be generated for each category of networks under the four configurations, which is 40. There are two categories of networks (i.e., SW and BA), so the size of B3 is $2 \times 10 \times 4 = 80$.

datasets	Name	Graphs	Categories	Avg. nodes	Avg. edges	Domain
B4	MUTAG	188	2	17.93	19.79	small molecules
B5	highschool_ct1	180	2	52.32	544.81	social networks
B6	bin_sulc	40	2	400	15960	brain networks

Table 2: **Description of real-world datasets in specific domains for binary classification.** Each comes from a specific domain and is further subdivided into different categories that belong to this domain. For instance, B4 is a dataset about small molecules named MUTAG. It contains 188 graphs with two categories, and the average nodes and average edges are 17.93 and 19.97, respectively.

datasets	Combinations
B7	B4, B5
B8	B5, B6

Table 3: **Description of combinations datasets for binary classification.** For instance, B4 and B5 are combined together to construct B7 for binary graph classification. The graphs in B4 are unified into one category (i.e., small molecules), and those in B5 belong to another (i.e., social networks).

B.2 Datasets for multi-class classification

datasets	Sample size	Node-scale configuration	Sparsity configuration	Total size
M1	10	[100]	[0.1]	3×10×1=30
M2	10	[100]	[0.1,0.9]	3×10×2=60
M3	10	[100,200]	[0.1,0.9]	3×10×4=120

Table 4: **Description of simulation datasets for multi-class classification.** ER, SW, and BA are used to construct simulation datasets for multi-class classification.

datasets	Name	Graphs	Categories	Avg. nodes	Avg. edges	Domain
M4	MSRC_9	221	8	40.58	97.94	computer vision
M5	MSRC_21	563	20	77.52	198.32	computer vision
M6	MSRC_21C	209	20	40.28	96.60	computer vision

Table 5: **Description of real-world datasets in specific domains for muiti-class classification.** Computer vision is a popular application domain of multi-class graph classification.

datasets	Combinations
M7	B4, B5, M4
M8	B4, B5, B6, M4, M5, M6

Table 6: **Description of combinations datasets for multi-class classification.** Within M8, M4-M6 are unified into the same category because the graphs therein come from the computer vision domain. Thus, M8 has four categories.

B.3 Hyperparameters of models

In Table 7, we provide the detailed parameters for each baseline model.

Models	hyperparameters
DHC-E	**none**
Graph2vec	wl_iterations = 2, dimension = 128, epochs = 10, learning_rate = 0.025, down_sampling = 0.0001, min_count = 5, seed = 42
IGE	feature_embedding_dimensions = [3,5], spectral_embedding_dimensions = [10,20], histogram_bins = [10,20], seed = 42
GL2vec	wl_iterations = 2, dimension = 128, epoches = 10, learning_rate = 0.025, down_sampling = 0.0001, min_count = 5, seed = 42

Table 7: **Hyperparameters of models.** The hyperparameters of baselines adopted in experiments follow the default values in the python package named *karate club*. Distinctively, DHC-E requires zero hyperparameters for its implementation.

References

[Albert *et al.*, 1999] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999.

[Di Martino *et al.*, 2017] Adriana Di Martino, David O'connor, Bosi Chen, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Joshua H Balsters, Leslie Baxter, Anita Beggiato, and Sylvie Bernaerts. Enhancing studies of the connectome in autism using the autism brain imaging data exchange ii. *Scientific Data*, 4(1):1–15, 2017.

[Erdős and Rényi, 1960] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[Faloutsos *et al.*, 1999] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. *ACM SIGCOMM Computer Communication Review*, 29(4):251–262, 1999.

[Li *et al.*, 2021] Yinzhi Li, Ningkai Wang, Hao Wang, Yating Lv, Qihong Zou, and Jinhui Wang. Surface-based single-subject morphological brain networks: Effects of morphological index, brain parcellation and similarity measure, sample size-varying stability and test-retest reliability. *NeuroImage*, 235:118018, 2021.

[Morris *et al.*, 2020] Christopher Morris, Nils M. Kriege, Franka Bause, Kristian Kersting, Petra Mutzel, and Marion Neumann. Tudataset: A collection of benchmark datasets for learning with graphs. In *ICML 2020 Workshop on Graph Representation Learning and Beyond (GRL+ 2020)*, 2020.

[Schaefer *et al.*, 2018] A. Schaefer, R. Kong, E. M. Gordon, T. O. Laumann, X. N. Zuo, A. J. Holmes, S. B. Eickhoff, and B. T. T. Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cereb Cortex*, 28(9):3095–3114, 2018.

[Wang *et al.*, 2016] H. Wang, X. Jin, Y. Zhang, and J. Wang. Single-subject morphological brain networks: connectivity mapping, topological characterization and test-retest reliability. *Brain Behav*, 6(4):e00448, 2016.

[Wu *et al.*, 2018] Huijun Wu, Hao Wang, and Linyuan Lü. Individual t1-weighted/t2-weighted ratio brain networks: Small-worldness, hubs and modular organization. *International Journal of Modern Physics C*, 29(05):1840007, 2018.