A FULL STATISTICS OF SHGB DATASETS

Table 4: Statistics of all 23 semi-hypergraph datasets in SHGB.

Name	#Nodes	#Edges	#Hyperedges	Avg. Node Degree	Avg. Hyperedge Degree	#Node Features	#Classes
MUSAE-Github	37,700	578,006	223,672	30.66	4.591	4,005 or 128	4
MUSAE-Facebook	22,470	342,004	236,663	30.44	9.905	4,714 or 128	4
MUSAE-Twitch-DE	9,498	306,276	297,315	64.49	7.661	3,170 or 128	2
MUSAE-Twitch-EN	7,126	70,648	13,248	19.83	3.666	3,170 or 128	2
MUSAE-Twitch-ES	4,648	118,764	77,135	51.10	5.826	3,170 or 128	2
MUSAE-Twitch-FR	6,549	225,332	172,653	68.81	5.920	3,170 or 128	2
MUSAE-Twitch-PT	1,912	62,598	74,830	65.48	7.933	3,170 or 128	2
MUSAE-Twitch-RU	4,385	74,608	25,673	34.03	4.813	3,170 or 128	2
MUSAE-Wiki-Chameleon	2,277	62,742	14,650	55.11	7.744	3,132 or 128	Regression
MUSAE-Wiki-Crocodile	11,631	341,546	121,431	58.73	4.761	13,183 or 128	Regression
MUSAE-Wiki-Squirrel	5,201	396,706	220,678	152.55	30.735	3,148 or 128	Regression
GRAND-ArteryAorta	5,848	5,823	11,368	1.991	1.277	4,651	3
GRAND-ArteryCoronary	5,755	5,722	11,222	1.989	1.273	4,651	3
GRAND-Breast	5,921	5,910	11,400	1.996	1.281	4,651	3
GRAND-Brain	6,196	6,245	11,878	2.016	1.296	4,651	3
GRAND-Lung	6,119	6,160	11,760	2.013	1.291	4,651	3
GRAND-Stomach	5,745	5,694	11,201	1.982	1.274	4,651	3
GRAND-Leukemia	4,651	6,362	7,812	2.736	1.324	4,651	3
GRAND-Lungcancer	4,896	6,995	8,179	2.857	1.334	4,651	3
GRAND-Stomachcancer	4,518	6,051	7,611	2.679	1.312	4,651	3
GRAND-KidneyCancer	4,319	5,599	7,369	2.593	1.297	4,651	3
Amazon-Computers	10,226	55,324	10,226	10.82	3.000	1,000	10
Amazon-Photos	6,777	45,306	6,777	13.37	4.800	1,000	10

B EXPERIMENTAL DETAILS

B.1 TRAINING DETAILS

We run all the experiments on NVIDIA A100 PCIe GPU with 40GB RAM (Sulis) and NVIDIA V100 NVLink GPU with 32GB RAM (JADE), with each experiment taking less than 2 minutes. Adam (Kingma & Ba, 2015) is used as the optimiser, and CosineAnnealingLR (Gotmare et al., 2019) is used as the learning rate scheduler for all training. All models are trained for 50 epochs. For each experiment, the nodes of the used semi-hypergraph are split into the train, validation, and test sets with a split ratio of 6:2:2. For node classification tasks, BCEWithLogitsLoss is used as the loss function, which is defined as:

$$\mathcal{L}_{\text{BCEWithLogits}}(\mathbf{y}, \hat{\mathbf{y}}) = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \cdot \log(\sigma(\hat{y}_i)) + (1 - y_i) \cdot \log(1 - \sigma(\hat{y}_i))\right] \tag{2}$$

where n is the total number of elements in $\mathbf{y}$ and $\hat{\mathbf{y}}$, y_i is the i-th element of $\mathbf{y}$, the batch of true values, and $\hat{y}_i$ is the i-th element of $\hat{\mathbf{y}}$, the batch of raw (i.e., non-sigmoid-transformed) predicted values. σ denotes the sigmoid function, which transforms the raw predictions into the range (0, 1). For node regression tasks, MSELoss is used as the loss function, which is defined as:

$$\mathcal{L}_{\text{MSE}}(\mathbf{y}, \hat{\mathbf{y}}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

where n is the total number of elements in $\mathbf{y}$ and $\hat{\mathbf{y}}$, y_i is the i-th element of $\mathbf{y}$, the batch of true values, and $\hat{y}_i$ is the i-th element of $\hat{\mathbf{y}}$, the batch of predicted values.

B.2 Hyperparameter Settings

We perform a hyperparameter search for the learning rate and keep the hidden layer dimension the same for different models, the hyperparameters used for training each architecture are listed in Table 5. All seven GNNs (GCN, GraphSAGE, GAT, GATv2, HyperConv, HyperAtten, and GraphSAINT) share the same learning rate, hidden dimension, and dropout rate. HyperAtten has an additional hyperparameter, which is the hyperedge aggregation function. This function determines how the hyperedge is constructed from the nodes within it. The possible options for this function are 'sum' and 'concatenate'. In this work, we have selected 'sum' as the hyperedge aggregation function. For GraphSAINT, the additional hyperparameters are subgraph size, measured by the number of nodes in the subgraph, and the batch size, which is the number of subgraphs to sample in each epoch. Different subgraph sizes are applied according to the sizes of the original hypergraphs.

Table 5: Hyperparameter selections for the experiments.

Method	Hyperparameter	Value
All	Learning rate	0.01
	Hidden dimension	32
	Dropout rate	0.5
HyperAtten	Hyperedge aggregation function	Sum
GraphSAINT	Subgraph size (MUSAE-GitHub, MUSAE-Facebook)	5000
	Subgraph size (MUSAE-Twitch-PT)	1000
	Subgraph size (others)	3000
	Batch size	5

C Results

We evaluate the performance of seven GNNs on all 23 SHGB datasets. Each experiment is repeated five times with different random seeds, and the results are summarised in Tables 6 to 11. In the node classification tasks of MUSAE, GCN and GraphSAGE perform the best in the on GitHub and Facebook, as shown in Table 6, while all GNNs perform roughly the same on Twitch, as shown in Table 7. In the three node regression tasks on MUSAE-Wiki, GraphSAINT stands out among other GNNs, as shown in Table 11. Table 8 shows that HyperConv and HyperAtten outperform other simple graph GNNs on five of the six hypergraphs in GRAND-Tissues. For hypergraphs in GRAND-Diseases as shown in Table 9, GraphSAGE exhibits a superior performance. For the two Amazon hypergraph datasets, as shown in Table 10, GraphSAINT consistently achieves the best performance. Overall, hypergraph GNNs tend to outperform simple graph GNNs on GRAND-Tissues and Amazon, perform equally as simple graph GNNs on MUSAE-Twitch, MUSAE-Wiki and GRAND-Diseases, and underperform simple graph GNNs on MUSAE-GitHub and MUSAE-Facebook.

Figure 7 summarises the performance of HyperConv (accuracy for prediction tasks, and MSE for regression tasks) with the three HypergraphSAINT samplers on 21 SHGB graphs other than MUSAE-GitHUB and MUSAE-Facebook, which are illustrated in Figure 5. The HypergraphSAINT sampling techniques generally enhance HyperConv's accuracy across most graphs, and are especially significant for regression tasks like MUSAE-Chameleon, MUSAE-Crocodile, and MUSAE-Squirrel.

Tables 12 to 17 reports the performances of LP-GNNs, GCN, GAT, and HyperConv on all 23 SHGB datasets. Notably, LP-GAT+HyperConv and LP-GCN+HyperConv surpass the other four methods in 18 of the 23 graphs. These results underscore the benefits of multi-level information integration in hypergraph representation learning.

Table 6: Accuracies of the selected GNNs on MUSAE-Facebook and GitHub datasets.

Method	Facebook	GitHub
RandomGuess	0.250	0.250
GCN	0.886 ± 0.001	$\mathbf{0.872 \pm 0.000}$
GraphSAGE	$\mathbf{0.902 \pm 0.002}$	0.871 ± 0.002
GAT	0.876 ± 0.001	0.864 ± 0.001
GATv2	0.901 ± 0.001	0.866 ± 0.001
HyperConv	0.792 ± 0.001	0.808 ± 0.001
HyperAtten	0.523 ± 0.002	0.775 ± 0.001
GraphSAINT	0.896 ± 0.001	0.871 ± 0.001

Table 7: Accuracies of the selected GNNs on the MUSAE-Twitch datasets.

Method	TwitchES	TwitchFR	TwitchDE	TwitchEN	TwitchPT	TwitchRU
RandomGuess	0.500	0.500	0.500	0.500	0.500	0.500
GCN	$\mathbf{0.721 \pm 0.004}$	0.624 ± 0.001	0.655 ± 0.002	$\mathbf{0.620 \pm 0.003}$	0.689 ± 0.006	0.745 ± 0.000
GraphSAGE	0.690 ± 0.002	0.616 ± 0.003	$\mathbf{0.657 \pm 0.001}$	0.605 ± 0.000	0.672 ± 0.013	0.745 ± 0.001
GAT	0.694 ± 0.002	0.623 ± 0.000	0.645 ± 0.004	0.594 ± 0.006	0.664 ± 0.007	0.743 ± 0.002
GATv2	0.710 ± 0.003	$\mathbf{0.625 \pm 0.001}$	0.651 ± 0.003	0.618 ± 0.005	0.687 ± 0.009	0.745 ± 0.000
HyperConv	0.715 ± 0.001	0.624 ± 0.002	0.654 ± 0.002	0.587 ± 0.007	$\mathbf{0.701 \pm 0.005}$	0.741 ± 0.001
HyperAtten	0.695 ± 0.000	0.623 ± 0.001	0.610 ± 0.003	0.553 ± 0.003	0.641 ± 0.000	0.743 ± 0.000
GraphSAINT	0.713 ± 0.008	0.622 ± 0.003	0.653 ± 0.004	0.610 ± 0.011	0.677 ± 0.006	$\mathbf{0.746 \pm 0.002}$

Table 8: Accuracies of the selected GNNs on the GRAND-Tissues datasets.

Method	ArteryAorta	ArteryCoronary	Breast	Brain	Lung	Stomach
RandomGuess	0.333	0.333	0.333	0.333	0.333	0.333
GCN	0.627 ± 0.007	0.662 ± 0.001	0.639 ± 0.010	0.625 ± 0.000	0.650 ± 0.000	$\mathbf{0.643 \pm 0.000}$
GraphSAGE	0.628 ± 0.002	0.663 ± 0.001	0.644 ± 0.000	0.618 ± 0.002	0.646 ± 0.005	0.630 ± 0.010
GAT	0.628 ± 0.004	0.663 ± 0.000	0.643 ± 0.001	0.625 ± 0.001	0.648 ± 0.004	0.643 ± 0.000
GATv2	0.630 ± 0.000	0.663 ± 0.000	0.644 ± 0.000	0.624 ± 0.001	0.650 ± 0.001	0.642 ± 0.000
HyperConv	0.626 ± 0.008	0.662 ± 0.000	$\mathbf{0.645 \pm 0.001}$	0.625 ± 0.000	0.650 ± 0.000	0.643 ± 0.000
HyperAtten	$\mathbf{0.647 \pm 0.003}$	$\mathbf{0.670 \pm 0.003}$	0.633 ± 0.003	$\mathbf{0.632 \pm 0.003}$	$\mathbf{0.661 \pm 0.004}$	0.636 ± 0.004
GraphSAINT	0.630 ± 0.000	0.663 ± 0.000	0.644 ± 0.000	0.625 ± 0.000	0.650 ± 0.000	0.643 ± 0.000

Table 9: Accuracies of the selected GNNs on the GRAND-Diseases datasets.

Method	Leukemia	LungCancer	StomachCancer	KidneyCancer
RandomGuess	0.333	0.333	0.333	0.333
GCN	0.582 ± 0.001	0.596 ± 0.001	0.602 ± 0.007	0.581 ± 0.002
GraphSAGE	$\mathbf{0.604 \pm 0.016}$	$\mathbf{0.615 \pm 0.015}$	0.602 ± 0.016	$\mathbf{0.596 \pm 0.014}$
GAT	0.587 ± 0.005	0.596 ± 0.000	0.600 ± 0.007	0.581 ± 0.003
GATv2	0.583 ± 0.000	0.591 ± 0.005	0.596 ± 0.002	0.579 ± 0.006
HyperConv	0.586 ± 0.003	0.593 ± 0.003	0.596 ± 0.004	0.577 ± 0.006
HyperAtten	0.593 ± 0.008	0.608 ± 0.008	$\mathbf{0.604 \pm 0.022}$	0.578 ± 0.012
GraphSAINT	0.583 ± 0.000	0.595 ± 0.000	0.596 ± 0.000	0.582 ± 0.000

Table 10: Accuracies of the selected GNNs on the Amazon datasets.

Method	Computers	Photos
RandomGuess	0.100	0.100
GCN	0.756 ± 0.041	0.295 ± 0.017
GraphSAGE	0.582 ± 0.108	0.366 ± 0.061
GAT	0.742 ± 0.043	0.434 ± 0.074
GATv2	0.566 ± 0.046	0.420 ± 0.075
HyperConv	0.842 ± 0.020	0.337 ± 0.059
HyperAtten	0.663 ± 0.005	0.465 ± 0.033
GraphSAINT	**0.875 ± 0.020**	**0.512 ± 0.141**

Table 11: MSEs (↓) of the selected GNNs on the MUSAE-Wiki datasets.

Method	Chameleon	Squirrel	Crocodile
GCN	7.319 ± 0.000	8.761 ± 0.001	6.779 ± 0.005
GraphSAGE	6.945 ± 0.005	8.310 ± 0.003	6.380 ± 0.005
GAT	6.557 ± 0.154	8.093 ± 0.054	6.249 ± 0.261
GATv2	7.290 ± 0.019	8.600 ± 0.011	6.717 ± 0.005
HyperConv	7.230 ± 0.002	8.706 ± 0.000	6.712 ± 0.001
HyperAtten	7.451 ± 0.000	8.782 ± 0.000	6.942 ± 0.000
GraphSAINT	**5.165 ± 0.027**	**7.541 ± 0.023**	**4.898 ± 0.035**

Table 12: Accuracies LP-GNNs and other baselines on MUSAE-Facebook and GitHub.

Method	Facebook	GitHub
RandomGuess	0.250	0.250
GCN	0.886 ± 0.001	0.872 ± 0.000
GAT	0.876 ± 0.001	0.864 ± 0.001
HyperConv	0.792 ± 0.001	0.808 ± 0.001
LP-GCN+GAT	**0.910 ± 0.001**	0.867 ± 0.001
LP-GCN+HyperConv	0.898 ± 0.000	**0.872 ± 0.000**
LP-GAT+HyperConv	0.905 ± 0.000	0.860 ± 0.002

Table 13: Accuracies LP-GNNs and other baselines on MUSAE-Twitch.

Method	TwitchES	TwitchFR	TwitchDE	TwitchEN	TwitchPT	TwitchRU
RandomGuess	0.500	0.500	0.500	0.500	0.500	0.500
GCN	0.721 ± 0.004	0.624 ± 0.001	0.655 ± 0.002	**0.620 ± 0.003**	0.689 ± 0.006	**0.745 ± 0.000**
GAT	0.694 ± 0.002	0.623 ± 0.000	0.645 ± 0.004	0.594 ± 0.006	0.664 ± 0.007	0.743 ± 0.002
HyperConv	0.715 ± 0.001	0.624 ± 0.002	0.654 ± 0.002	0.587 ± 0.007	0.701 ± 0.005	0.741 ± 0.001
LP-GCN+GAT	0.727 ± 0.001	0.623 ± 0.001	**0.662 ± 0.001**	0.612 ± 0.002	0.686 ± 0.008	0.745 ± 0.001
LP-GCN+HyperConv	**0.729 ± 0.001**	**0.626 ± 0.000**	0.657 ± 0.001	0.607 ± 0.001	**0.696 ± 0.004**	0.744 ± 0.000
LP-GAT+HyperConv	0.714 ± 0.001	0.622 ± 0.000	0.654 ± 0.001	0.608 ± 0.003	0.672 ± 0.003	0.744 ± 0.000

Table 14: Accuracies of LP-GNNs and other baselines on GRAND-Tissues.

Method	ArteryAorta	ArteryCoronary	Breast	Brain	Lung	Stomach
RandomGuess	0.333	0.333	0.333	0.333	0.333	0.333
GCN	0.627 ± 0.007	0.662 ± 0.001	0.639 ± 0.010	0.625 ± 0.000	0.650 ± 0.000	0.643 ± 0.000
GAT	0.628 ± 0.004	0.663 ± 0.000	0.643 ± 0.001	0.625 ± 0.001	0.648 ± 0.004	0.643 ± 0.000
HyperConv	0.626 ± 0.008	0.662 ± 0.000	0.645 ± 0.001	0.625 ± 0.000	0.650 ± 0.000	0.643 ± 0.000
LP-GCN+GAT	0.627 ± 0.000	0.641 ± 0.000	0.626 ± 0.001	0.625 ± 0.003	0.627 ± 0.001	0.626 ± 0.001
LP-GCN+HyperConv	0.647 ± 0.003	0.660 ± 0.003	0.652 ± 0.006	0.637 ± 0.002	0.654 ± 0.004	0.654 ± 0.001
LP-GAT+HyperConv	**0.649** ± **0.001**	**0.664** ± **0.001**	**0.657** ± **0.001**	**0.645** ± **0.001**	**0.662** ± **0.002**	**0.656** ± **0.002**

Table 15: Accuracies of LP-GNNs and other baselines on GRAND-Diseases.

Method	Leukemia	LungCancer	StomachCancer	KidneyCancer
RandomGuess	0.333	0.333	0.333	0.333
GCN	0.582 ± 0.001	0.596 ± 0.001	0.602 ± 0.007	0.581 ± 0.002
GAT	0.587 ± 0.005	0.596 ± 0.000	0.600 ± 0.007	0.581 ± 0.003
HyperConv	0.586 ± 0.003	0.593 ± 0.003	0.596 ± 0.004	0.577 ± 0.006
LP-GCN+GAT	0.590 ± 0.002	0.583 ± 0.002	0.581 ± 0.004	0.584 ± 0.002
LP-GCN+HyperConv	**0.604** ± **0.004**	0.614 ± 0.006	0.605 ± 0.003	0.609 ± 0.004
LP-GAT+HyperConv	0.601 ± 0.002	**0.618** ± **0.003**	**0.621** ± **0.002**	**0.621** ± **0.002**

Table 16: Accuracies of LP-GNNs and other baselines on Amazon.

Method	Computers	Photos
RandomGuess	0.1	0.1
GCN	0.756 ± 0.041	0.295 ± 0.017
GAT	0.742 ± 0.043	0.434 ± 0.074
HyperConv	0.842 ± 0.020	0.337 ± 0.059
LP-GCN+GAT	**0.930** ± **0.000**	0.711 ± 0.005
LP-GCN+HyperConv	0.913 ± 0.001	0.698 ± 0.005
LP-GAT+HyperConv	**0.930** ± **0.007**	**0.715** ± **0.003**

Table 17: MSEs (↓) of LP-GNNs and other baselines on MUSAE-Wiki.

Method	Squirrel	Crocodile	Chameleon
GCN	7.319 ± 0.000	8.761 ± 0.001	6.779 ± 0.005
GAT	6.557 ± 0.154	8.093 ± 0.054	6.249 ± 0.261
HyperConv	7.230 ± 0.002	8.706 ± 0.000	6.712 ± 0.001
LP-GCN+GAT	7.554 ± 0.029	**4.827** ± **0.087**	5.203 ± 0.147
LP-GCN+HyperConv	7.313 ± 0.007	4.851 ± 0.014	5.515 ± 0.008
LP-GAT+HyperConv	**6.049** ± **0.204**	4.875 ± 0.031	**4.665** ± **0.050**

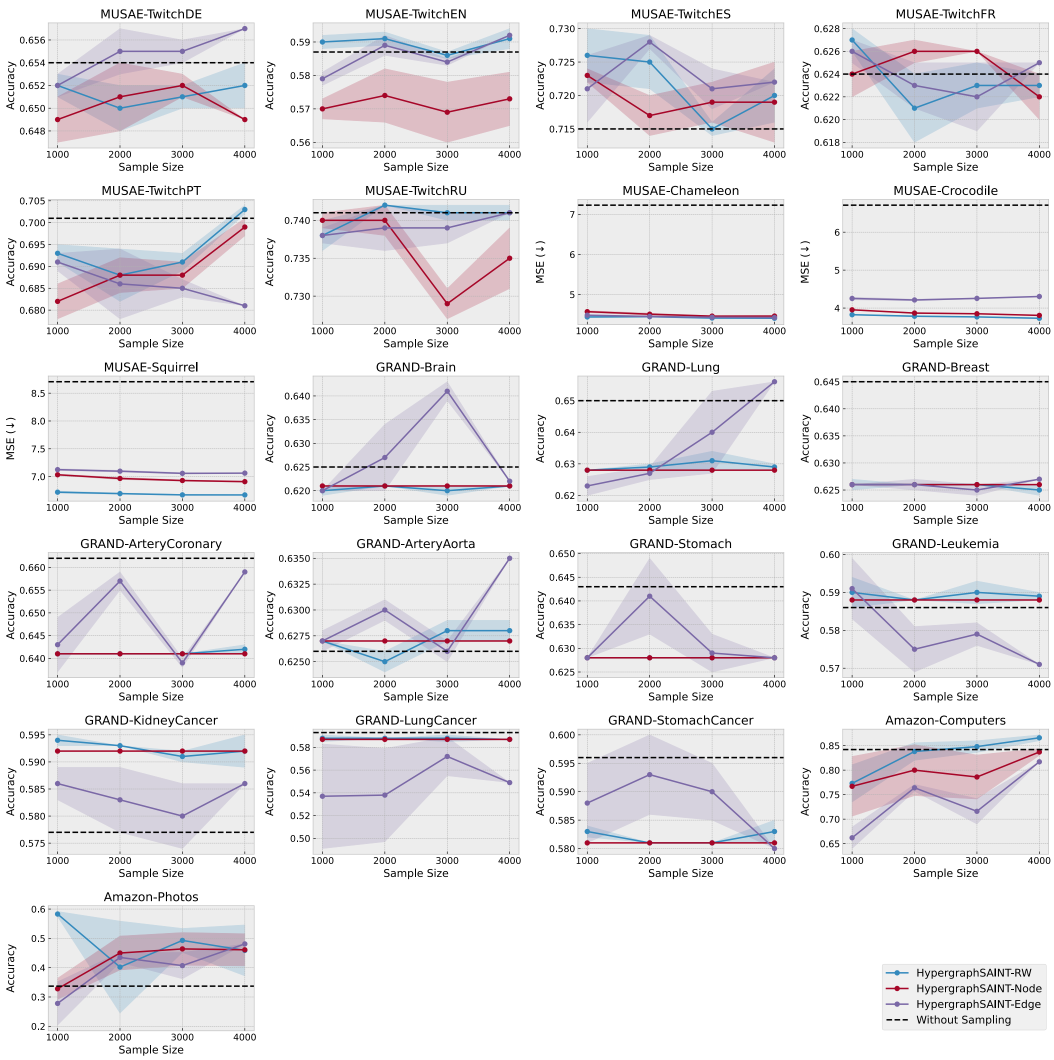

Figure 7: Node prediction performances of different sampling techniques on 21 SHGB datasets.

D Data Accessibility

The source code and full datasets of SHGB is publicly available at `https://anonymous-url/`. While the raw dataset in JSON format is hosted at `https://anonymous-url/`, we recommend the users to access the datasets through our Python library `anonymous-library`, which is installable via `pip`. This would allow the users to read the semi-hypergraphs in the format of PyTorch Geometric `Data` objects.

E Licence

The raw data for the MUSAE datasets are licenced under the the GNU General Public Licence, version 3 (GPLv3)[3]. The raw data for the GRAND datasets are licenced under the Creative Commons Attribution-ShareAlike 4.0 International Public Licence (CC BY-SA 4.0)[4]. The raw data for the Amazon datasets are licenced under the Amazon Service licence[5]. Having carefully observed the licence requirements of all data sources and code dependencies, we apply the following licence to our source code and datasets:

- The source code of SHGB is licenced under the MIT licence[6];
- The MUSAE and GRAND datasets are licenced under the GPLv3 licence[3];
- The Amazon datasets are licenced under the Amazon Service licence[5].

F Ethics Statement

All datasets constructed in SHGB are generated from public open-source datasets, and the original raw data downloaded from the data sources do not contain any personally identifiable information or other sensitive contents. The node prediction tasks for the SHGB datasets are designed to ensure that they do not, by any means, lead to discriminations against any social groups. Therefore, we are not aware of any social or ethical concern of SHGB. Since SHGB is a general benchmarking tool for representation learning on complex graphs, we also do not forsee any direct application of SHGB to malicious purposes. However, the users of SHGB should be aware of any potential negative social and ethical impacts that may arise from their chosen downstream datasets or tasks outside of SHGB, if they intend to use the SHGB datasets as pre-training datasets to perform trasnfer learning.

[3] `https://www.gnu.org/licenses/gpl-3.0.html`
[4] `https://creativecommons.org/licenses/by-sa/4.0/`
[5] `https://s3.amazonaws.com/amazon-reviews-pds/LICENSE.txt`
[6] `https://opensource.org/license/mit/`