

Appendix

A.1 Dataset Statistics

Eight publicly available graph datasets are used to evaluate the proposed KDGA framework. An overview summary of the statistical characteristics of datasets is given in Tab. A1.

Cora, Citeseer, and Pubmed. These three are citation network benchmark datasets for node classification. In these datasets, nodes represent papers, and edges denote citations of one paper by another. Node features are the bag-of-words representation of papers, and the node label is the academic topic of a paper. The data splitting of these dataset is the same as [22].

Cornell, Texas, and Wisconsin. Cornell, Texas, and Wisconsin¹ are three sub-datasets of WebKB, which is a webpage dataset collected from computer science departments of various universities by Carnegie Mellon University. In these datasets, nodes represent web pages, and edges are hyperlinks between them. Node features are the bag-of-words representation of web pages. The nodes are manually classified into five categories: student, project, course, staff, and faculty.

Actor. This dataset is a subgraph of the film-director-actor-writer network. In this dataset, nodes represent actors, and edges are their co-occurrence on the same Wikipedia page. Node features are some keywords in the Wikipedia pages. The nodes are manually classified into five categories in terms of the words of the actor’s Wikipedia.

Chameleon and Squirrel. Chameleon and Squirrel are two page-page networks on specific topics in Wikipedia. In these datasets, nodes represent web pages, and edges are mutual links between pages. Node features are several informative nouns in the Wikipedia pages. The nodes are classified into five categories in terms of the number of the average monthly traffic of the web page.

Table A1: Statistical information of the datasets.

Dataset	Cora	Citeseer	Chameleon	Squirrel	Texas	Cornell	Wisconsin	Actor
# Nodes	2708	3327	2277	5210	183	183	251	7600
# Edges	5278	4614	3142	198493	279	277	450	26659
# Features	1433	3703	2325	2089	1703	1703	1703	932
# Classes	7	6	5	5	5	5	5	5
Homophily ratio r	0.81	0.74	0.23	0.22	0.11	0.30	0.21	0.22
Label Rate	5.2%	3.6%	48%	48%	48%	48%	48%	48%

A.2 Hyperparameters and Search Space

All baselines and our approach are implemented based on the standard implementation in the DGL library [42] using the PyTorch 1.6.0 library with Intel(R) Xeon(R) Gold 6240R @ 2.40GHz CPU and NVIDIA V100 GPU. The following hyperparameters are set for all datasets: weight decay $decay = 5e-4$; Maximum Epoch $E = 500$; Layer number $L = 2$, sampling temperature $\tau_2 = 1.0$. The other dataset-specific hyperparameters are determined by a hyperparameter search tool - NNI for each dataset, including hidden dimension F , learning rate lr , loss weight λ and κ , fusion factor α , and distillation temperature τ_1 . The hyperparameter search space is shown in Tab. A2, and the model with the highest validation accuracy is selected for testing. The best hyperparameter choices are available in the supplementary.

Table A2: Hyperparameter search space.

Hyperparameters	Search Space
Hidden Dimension F	[64, 128, 256]
Learning Rate lr	[1e-2, 5e-3, 1e-3]
Loss Weight λ	[0.1, 0.5, 1.0]
Loss Weight κ	[0.1, 0.5, 1.0, 5.0, 20.0]
Fusion Factor α	[0.1, 0.3, 0.5, 1.0]
Temperature τ_1	[1.0, 1.1, 1.2, 1.3, 1.4]

¹Cornell, Texas, and Wisconsin are three sub-datasets of WebKB1 from <http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-11/www/wskb>.