Dataset/Model	GCN	GIN	GAT	GraphSAGE
IMDB-BINARY	0.558/0.573/0.830/**0.833**	0.563/0.525/0.788/**0.825**	0.612/0.643/0.817/**0.825**	0.582/0.550/**0.822**/0.813
IMDB-MULTI	0.408/0.414/0.417/**0.537**	0.357/**0.363**/0.318/0.327	0.354/0.400/0.426/**0.524**	0.372/0.423/0.388/**0.497**
MSRC_21	0.614/0.746/0.798/**0.860**	0.412/0.798/0.827/**0.842**	0.497/0.819/0.789/**0.833**	0.418/0.743/0.789/**0.813**

Table 1. **Additional covariant shift/GNN backbones**: **Graph size** as covariate shift (a new type of covariate shift with additional GNN backbones). For each entry, the four numbers correspond to different data selection methods: **Random/KIDD/LAVA/Ours**. For ease of demonstration, we omit the GNN training on full dataset and only report the performance (mean accuracy) with 20% selected training data. The best performances are marked in **bold** and the second best performances are underlined.

Dataset/Model	GCN	GIN	GAT	GraphSAGE
IMDB-BINARY	×	×	0.662/0.813/0.835/**0.858**	0.738/0.797/0.835/**0.855**
IMDB-MULTI	×	×	0.067/0.133/0.790/**0.800**	0.132/0.137/0.570/**0.580**
MSRC_21	×	×	0.713/0.740/0.842/**0.857**	0.731/0.711/0.827/**0.842**

Table 2. **Additional covariant shift/GNN backbones**: **Graph density** as covariate shift (with additional GNN backbones). For each entry, the four numbers correspond to different data selection methods: **Random/KIDD/LAVA/Ours**. For ease of demonstration, we omit the GNN training on full dataset and only report the performance (mean accuracy) with 20% selected training data. The best performances are marked in **bold** and the second best performances are underlined. The repeated experiment settings will be marked in × (i.e. already included in the original paper).

	Procedure / Dataset	IMDB-BINARY	IMDB-MULTI	MSRC_21
Off-line Computation	FGW Pairwise distance	7.41	9.61	18.18
	Label-informed pairwise distance	0.04	0.06	0.24
On-line Computation	GREAT algorithm (Algorithm 2)	0.28	0.52	0.11
	LAVA algorithm	0.09	0.14	0.03
GNN Training Time	GCN (w/ 10% data)	13.45	16.36	9.59
	GCN (w/ 20% data)	17.64	21.40	13.85
	GCN (w/ 50% data)	29.92	45.82	19.57

Table 3. **Empirical run-time behavior (in seconds)**. We can observe that the off-line procedures can be run comparable to a single GNN training time and the on-line procedure has a negligible runtime compared to GNN training time. In addition, we can achieve significantly better performance compared to LAVA with slight additional on-line runtime.

Dataset/Model	GCN	GIN	GAT	GraphSAGE
ogbg-molbace	**0.555**/0.492/0.465/0.535	0.450/0.579/0.577/**0.644**	0.477/0.477/**0.502**/0.492	0.522/0.496/0.544/**0.629**
ogbg-molbbbp	0.670/0.659/0.657/**0.672**	0.670/0.470/0.864/**0.884**	0.623/0.581/0.651/**0.713**	0.628/0.576/0.610/**0.637**
ogbg-molhiv	0.570/0.522/**0.631**/0.622	0.640/0.601/0.755/**0.760**	0.557/0.554/**0.708**/0.644	0.489/0.572/**0.686**/0.622

Table 4. **Additional datasets/covariant shift/GNN backbones**: **Graph size** as covariate shift (a new type of covariate shift with additional GNN backbones). For each entry, the four numbers correspond to different data selection methods: **Random/KIDD/LAVA/Ours**. For ease of demonstration, we omit the GNN training on full dataset and only report the performance (mean AUCROC) with 20% selected training data. We train each model for 100 epochs. The best performances are marked in **bold** and the second best performances are underlined.

Dataset/Model	GCN	GIN	GAT	GraphSAGE
ogbg-molbace	0.529/0.511/0.541/**0.585**	0.622/0.631/0.655/**0.672**	0.472/**0.575**/0.515/0.518	0.459/0.530/0.514/**0.536**
ogbg-molbbbp	0.528/0.453/0.503/**0.571**	0.480/0.469/**0.675**/0.644	0.486/0.414/0.511/**0.538**	0.472/0.424/0.491/**0.533**
ogbg-molhiv	0.598/0.541/**0.627**/0.611	0.590/0.607/**0.746**/0.728	0.593/0.608/**0.644**/0.639	0.602/0.579/0.635/**0.638**

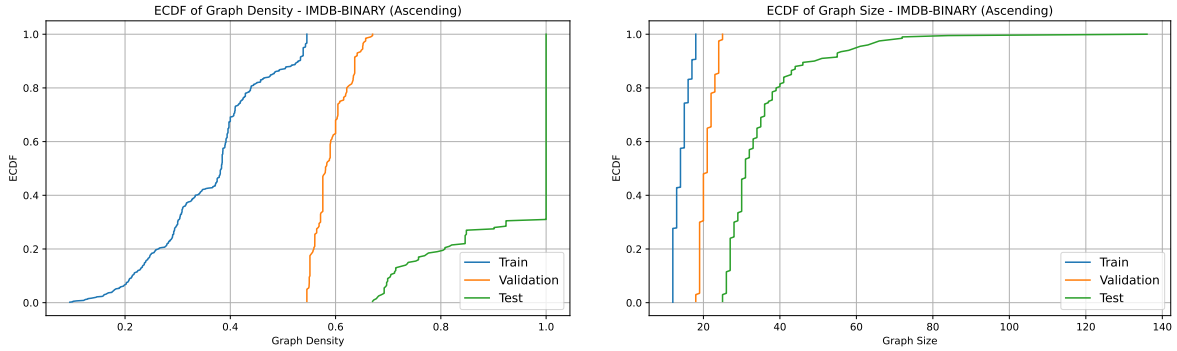
Table 5. **Additional datasets/GNN backbones**: **Graph density** as covariate shift (with additional GNN backbones). For each entry, the four numbers correspond to different data selection methods: **Random/KIDD/LAVA/Ours**. For ease of demonstration, we omit the GNN training on full dataset and only report the performance (mean AUCROC) with 20% selected training data. We train each model for 100 epochs. The best performances are marked in **bold** and the second best performances are underlined.

Dataset/Selection Method	Top τ-fraction densest	Random selection	Our method
IMDB-BINARY	0.590/0.795/0.868	0.737/0.660/0.827	**0.805/0.855/0.890**
IMDB-MULTI	0.439/**0.450**/0.467	0.139/0.092/0.080	**0.588**/0.349/**0.603**
MSRC_21	0.529/0.675/0.865	0.576/0.702/0.830	**0.681/0.789/0.906**

Table 6. **Comparison to top-τ selection and random selection**. Each entry shows accuracy from training a GCN with 10%/20%/50% of the training data. The best performances are marked in **bold** and the second best performances are underlined.

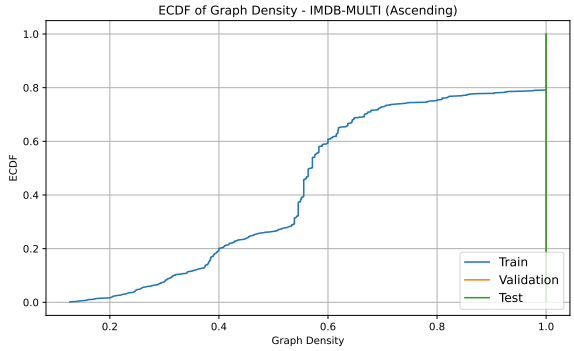
Type	Model	Data	IMDB-BINARY	IMDB-MULTI	MSRC_21	ogbg-molbace	ogbg-molbbbp	ogbg-molhiv
GDA	AdaGCN	Full	0.808± 0.015	0.073± 0.000	0.319± 0.032	0.641± 0.018	**0.764**± 0.009	0.471± 0.007
	GRADE	Full	0.822± 0.012	0.123± 0.061	0.804± 0.011	**0.685**± 0.034	0.497± 0.025	0.567± 0.004
	ASN	Full	0.782± 0.030	0.119± 0.047	0.833± 0.033	0.650± 0.034	0.511± 0.006	0.480± 0.093
	UDAGCN	Full	0.807± 0.013	0.114± 0.049	0.351± 0.019	0.541± 0.034	0.634± 0.013	0.451± 0.036
	A2GNN	Full	0.795± 0.023	0.114± 0.029	0.828± 0.039	0.569± 0.049	0.538± 0.022	0.623± 0.002
Vanilla	GCN	Random 20%	0.660± 0.012	0.092± 0.032	0.702± 0.045	0.529± 0.124	0.528± 0.030	0.598± 0.003
		LAVA 20%	0.620± 0.000	0.092± 0.032	0.819± 0.011	0.541± 0.067	0.503± 0.043	0.591± 0.030
		GENIUS 20%	0.830± 0.021	0.349± 0.323	0.797± 0.008	0.585± 0.074	0.571± 0.035	0.583± 0.006
	GIN	Random 20%	0.710± 0.049	0.180± 0.005	0.801± 0.046	0.622± 0.028	0.480± 0.041	0.590± 0.033
		LAVA 20%	0.778± 0.045	0.170± 0.009	0.851± 0.012	0.655± 0.067	0.644± 0.021	0.638± 0.012
		GENIUS 20%	0.832± 0.025	0.266± 0.133	**0.860**± 0.007	0.662± 0.006	0.665± 0.053	**0.644**± 0.017
	GAT	Random 20%	0.662± 0.029	0.067± 0.005	0.713± 0.008	0.472± 0.034	0.486± 0.041	0.593± 0.012
		LAVA 20%	0.835± 0.002	0.790± 0.002	0.842± 0.026	0.515± 0.019	0.511± 0.069	0.602± 0.017
		GENIUS 20%	**0.858**± 0.005	**0.800**± 0.133	0.857± 0.008	0.518± 0.026	0.538± 0.098	0.598± 0.004
	GraphSAGE	Random 20%	0.738± 0.059	0.132± 0.036	0.731± 0.027	0.459± 0.057	0.472± 0.016	0.602± 0.006
		LAVA 20%	0.835± 0.005	0.570± 0.292	0.827± 0.015	0.514± 0.132	0.491± 0.095	0.537± 0.067
		GENIUS 20%	0.855± 0.005	0.580± 0.281	0.842± 0.007	0.536± 0.062	0.533± 0.037	0.541± 0.014

Table 7. Performance comparison across GDA and vanilla methods for graph density shift. We use **bold**/underline to indicate the 1st/2nd best results. We add one GDA baseline, three more datasets and two GNN backbones. For the three OGB datasets, AUCROC is used for evaluation. Our method can achieve top-2 performance across datasets.

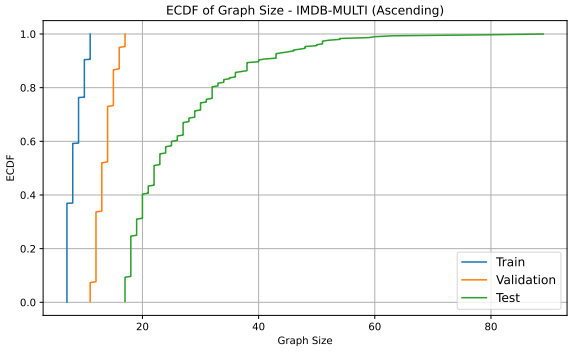


(a) IMDB-BINARY (Density)

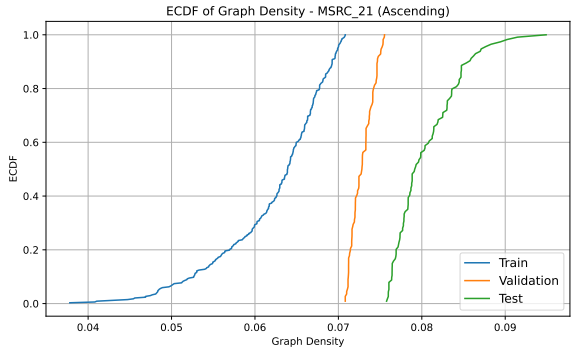
(b) IMDB-BINARY (Size)



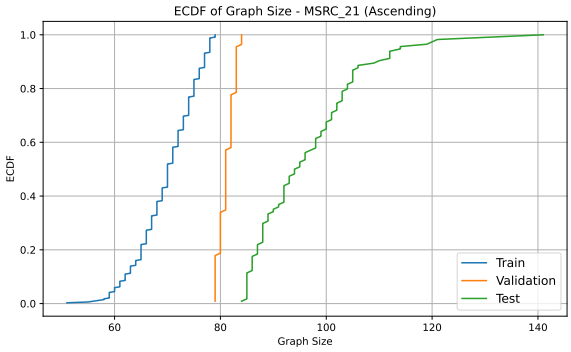
(c) IMDB-MULTI (Density)



(d) IMDB-MULTI (Size)

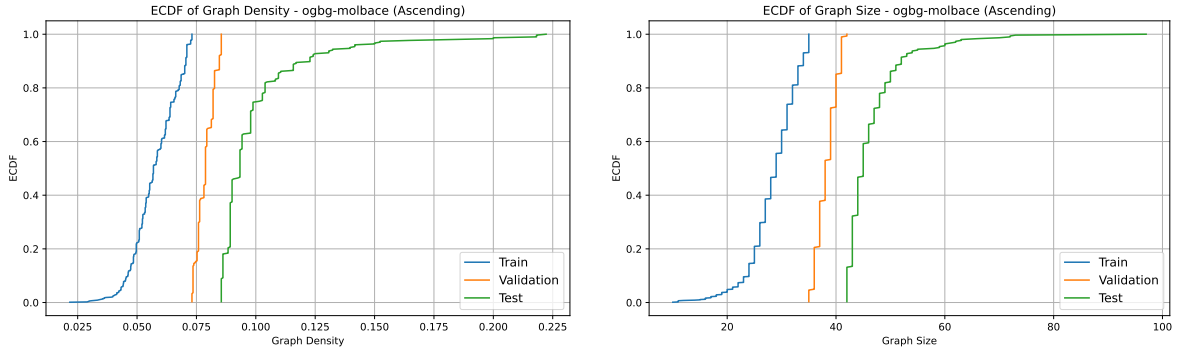


(e) MSRC_21 (Density)



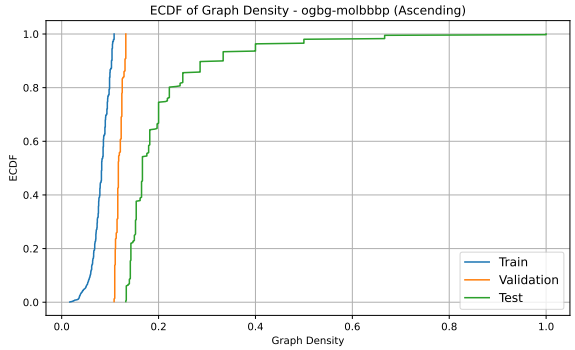
(f) MSRC_21 (Size)

Figure 1. **ECDF plots of graph density and size for IMDB-BINARY, IMDB-MULTI, and MSRC_21 datasets**. The Blue, Orange, and Green curves represent the distributions of the training, validation, and test splits, respectively. Graphs are sorted in ascending order by the specified shift (density or size).

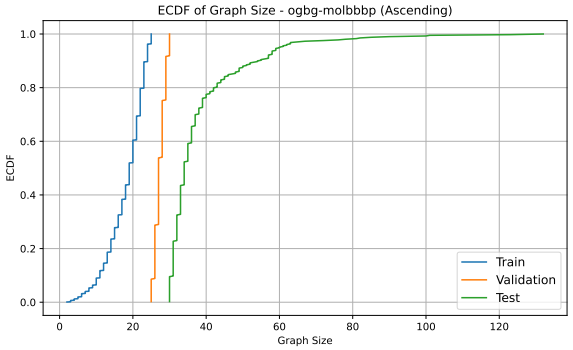


(a) `ogbg-molbace` (Density)

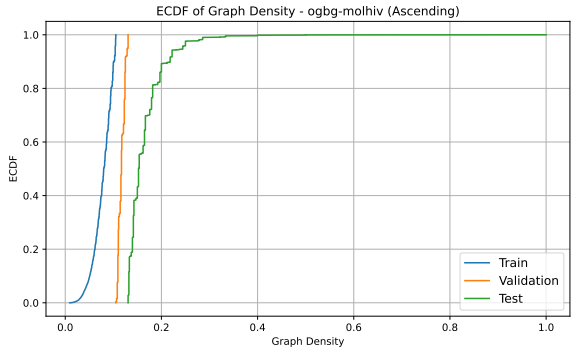
(b) `ogbg-molbace` (Size)



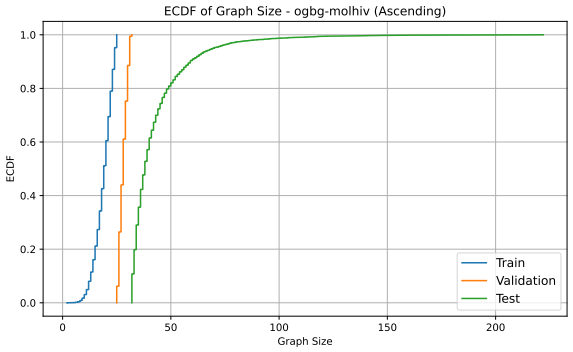
(c) `ogbg-molbbbp` (Density)



(d) `ogbg-molbbbp` (Size)



(e) `ogbg-molhiv` (Density)



(f) `ogbg-molhiv` (Size)

Figure 2. **ECDF plots of graph density and size for `ogbg-molbbbp`, `ogbg-molbace`, and `ogbg-molhiv` datasets**. The Blue, Orange, and Green curves represent the distributions of the training, validation, and test splits, respectively. Graphs are sorted in ascending order by the specified shift (density or size).