

Table 1: Ablation study of FAITH on transductive and inductive settings

Dataset ↓	Metric ↓	GLNN	w/o NDE	w/o OSM	FAITH
Transductive Setting					
Cora	Accuracy	81.50 \pm 2.23	83.20 \pm 0.31	83.62 \pm 1.78	83.96 \pm 0.47
	IF	101.23 \pm 2.47	10.28 \pm 2.47	182.27 \pm 38.65	12.06 \pm 1.30
Citeseer	Accuracy	71.46 \pm 3.16	72.67 \pm 0.29	73.02 \pm 0.32	74.10 \pm 0.60
	IF	5.68 \pm 2.36	0.65 \pm 0.23	8.97 \pm 0.53	1.78 \pm 1.57
Pubmed	Accuracy	77.90 \pm 3.35	79.54 \pm 0.59	80.16 \pm 0.49	81.72 \pm 0.33
	IF	413.38 \pm 42.92	46.38 \pm 3.09	367.19 \pm 62.73	47.96 \pm 3.38
Coauthor-CS	Accuracy	90.02 \pm 1.05	92.34 \pm 1.07	92.89 \pm 0.31	93.37 \pm 0.21
	IF	291962.86 \pm 13825.80	899.33 \pm 312.42	166746.63 \pm 88623.72	710.17 \pm 74.24
Inductive Setting					
Cora	Accuracy	72.64 \pm 0.39	73.50 \pm 0.43	72.83 \pm 0.41	73.62 \pm 0.41
	IF	4.63 \pm 0.03	2.06 \pm 0.04	4.44 \pm 0.06	2.03 \pm 0.05
Citeseer	Accuracy	61.02 \pm 0.69	70.76 \pm 0.88	69.63 \pm 0.92	71.00 \pm 0.86
	IF	1.69 \pm 0.09	0.45 \pm 0.05	0.84 \pm 0.83	0.33 \pm 0.05
Pubmed	Accuracy	77.68 \pm 0.64	79.13 \pm 0.17	79.23 \pm 0.24	80.44 \pm 0.17
	IF	0.44 \pm 0.08	0.28 \pm 0.02	0.43 \pm 0.03	0.29 \pm 0.02
Coauthor-CS	Accuracy	88.77 \pm 0.84	91.20 \pm 0.78	91.43 \pm 0.28	92.04 \pm 0.84
	IF	118.64 \pm 69.17	244.30 \pm 174.73	1733.49 \pm 1334.03	114.69 \pm 76.95

Table 2: Ablation study of FAITH★ on income dataset

Metric	GLNN	w/o NDE	w/o OSM	FAITH	FAITH★
Accuracy	85.56 \pm 2.21	84.92 \pm 1.41	85.23 \pm 2.96	84.55 \pm 1.03	85.08 \pm 0.31
IF	359.77 \pm 48.51	43.78 \pm 9.32	376.87 \pm 56.23	29.04 \pm 4.71	113.56 \pm 27.54
GD	1.17 \pm 0.25	1.36 \pm 2.21	1.23 \pm 0.12	1.55 \pm 0.08	1.40 \pm 0.02

Table 3: Comparison of different Dirichlet energy distillation approaches

Dataset ↓	Metric ↓	DE ratio distillation	Distill N2N	Distill Neighbourhood DE	FAITH
Cora	Accuracy	83.14 \pm 1.30	83.32 \pm 1.11	83.62 \pm 1.78	83.96 \pm 0.47
	IF	240.86 \pm 30.57	222.37 \pm 53.37	182.27 \pm 38.65	12.06 \pm 1.30
Citeseer	Accuracy	72.98 \pm 0.34	72.80 \pm 0.80	73.02 \pm 0.32	74.10 \pm 0.60
	IF	5.75 \pm 3.96	5.61 \pm 3.91	8.97 \pm 0.53	1.78 \pm 1.57
Pubmed	Accuracy	80.00 \pm 0.38	80.16 \pm 0.49	80.44 \pm 0.23	81.72 \pm 0.33
	IF	374.40 \pm 68.63	384.06 \pm 53.51	367.19 \pm 62.73	47.96 \pm 3.38
a-photo	Accuracy	91.06 \pm 2.56	91.10 \pm 1.27	91.34 \pm 2.57	92.48 \pm 0.03
	IF	24600.19 \pm 3884.21	20259.59 \pm 3634.08	19618.12 \pm 3119.82	1291.08 \pm 110.41

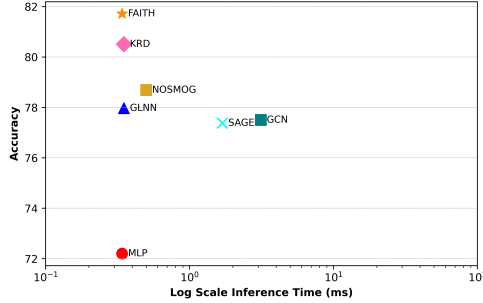
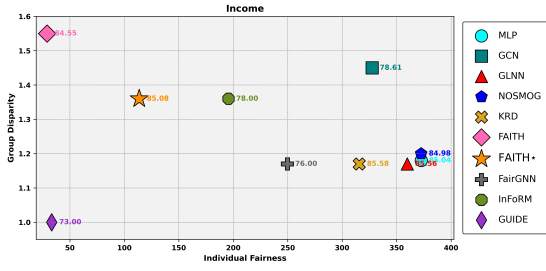


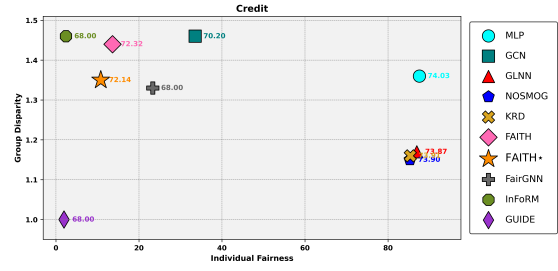
Figure 1: Utility vs. latency tradeoff on PubMed dataset.

Table 4: Comparison of Cut values of different methods.

Dataset	MLP	GLNN	NOSMOG	KRD	DE ratio distillation	FAITH
Cora	0.7050	0.8410	0.9370	0.9272	0.9353	0.9526
Citeseer	0.8313	0.9387	0.9527	0.9434	0.9391	0.9851
Pubmed	0.8144	0.9157	0.9501	0.9324	0.9137	0.9643
a-photo	0.6648	0.8756	0.8975	0.9109	0.8772	0.9278



(a) Income



(b) Credit

Figure 2: Tradeoff between individual and group fairness across methods on the Income and Credit datasets. An ideal method should lie in the bottom left corner with higher AUC value.

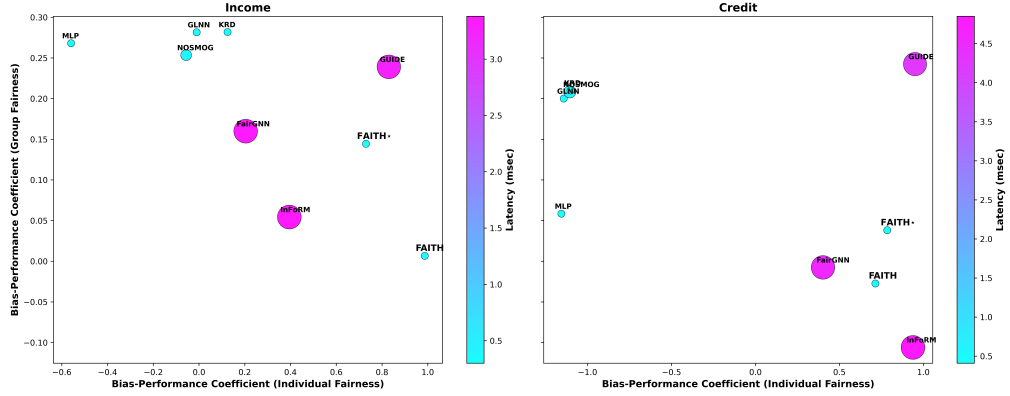


Figure 3: Utility, fairness and latency tradeoff between GNN-to-MLP and fair GNN baselines. An ideal method should lie in the top right corner with smaller circle. The radius of circle is proportional to the inference time.