Decoupled Graph Energy-based Model for Node Out-of-Distribution Detection on Heterophilic Graphs

ICLR 2025

Yuhan Chen[1]*, Yihong Luo[2]*, Yifan Song[2], Pengwen Dai[1], Jing Tang[2], Xiaochun Cao[1] (*Equally contributed)

[1]Sun Yat-sen University, [2]The Hong Kong University of Science and Technology

Background

- We focus on the **node Out-of-Distribution detection (OOD)** task. Extensive efforts have focus on image area (**i.i.d.**), but OOD detection on **nodes in graph learning (non-i.i.d.)** remains underexplored.
- Existing node OOD detection method are mostly based on **homophilic assumption**, will experience dramatic performance degradation when dealing with **heterophilic graphs.**

Methodology: Decoupled Graph Energy-based Model (DeGEM)

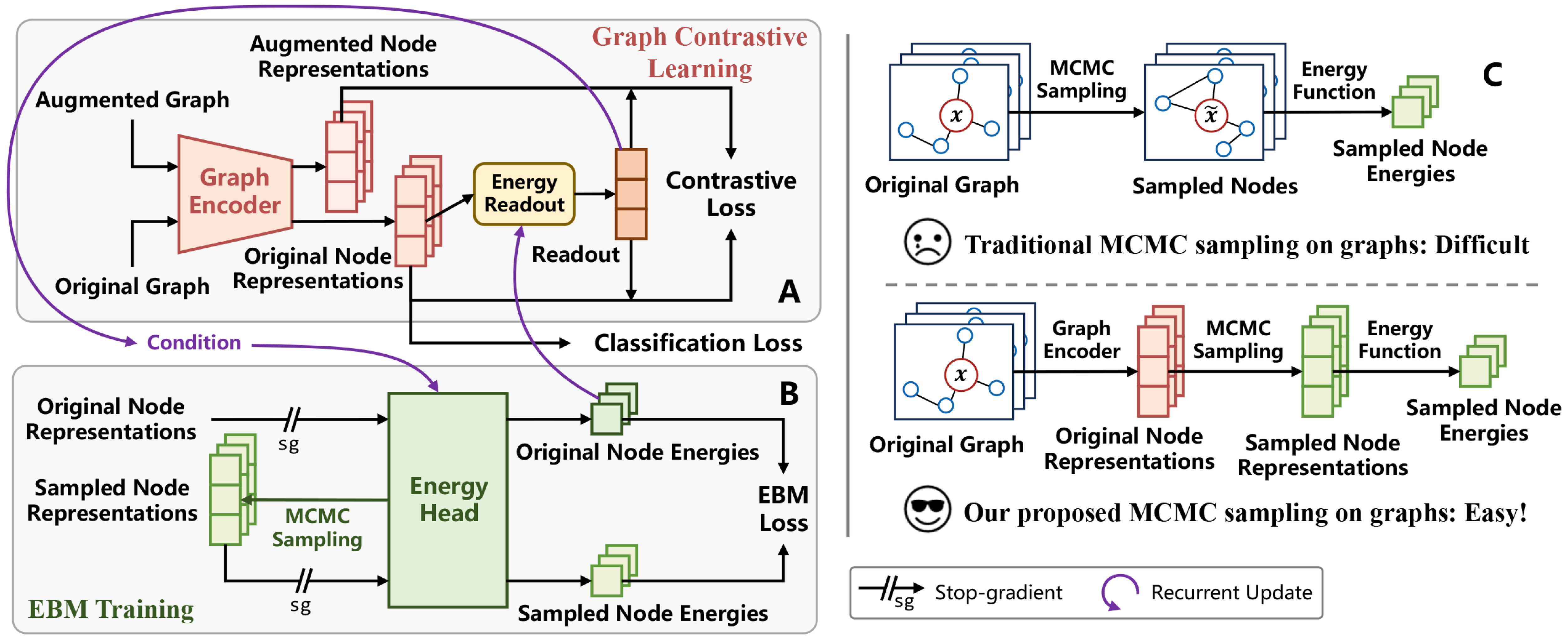


Fig 1. Framework of **DeGEM**

Motivation

We utilize **graph energy-based model (EBM)** in node OOD detection, and model the true node distribution $p_d(x)$ by Boltzmann distribution $p_\theta(x) = \frac{\exp(-E_\theta(x))}{Z_\theta}$, which can be trained by negative log-likelihood (NLL) loss:

$$\mathcal{L}_{\mathrm{NLL}} = -\mathbb{E}_{x\sim p_d(x)}[\log p_\theta(x)] \Leftrightarrow \mathbb{E}_{x\sim p_d(x)}[E_\theta(x)] - \mathbb{E}_{\tilde{x}\sim p_\theta(\tilde{x})}[E_\theta(\tilde{x})]$$

For efficiently training the graph EBM, we **decouple** the graph EBM into two parts: a graph encoder g_α and an energy head f_ω: $E_\theta(x) = f_\omega \circ g_\alpha(x, \mathcal{N}(x))$, such that the graph encoder first extract node features and topology information, then the energy-head only needs to sampling node representations in latent space.

Graph Contrastive Learning for Training Graph Encoder

We train a powerful graph encoder using **Graph Contrastive Learning (GCL).** Following DGI, given the original and augmented graph $\{\mathbf{X},\mathbf{A}\}$ & $\{\mathbf{X}',\mathbf{A}'\}$, we obtain the original and augmented node representations $\boldsymbol{h}_i = g_\alpha(x_i, \mathcal{N}(x_i))$ & $\boldsymbol{h}_i' = g_\alpha(x_i', \mathcal{N}(x_i'))$, and the readout summary $\boldsymbol{s} = \frac{1}{N}\sum_i^N \boldsymbol{h}_i$, using a discriminator D_ψ the **contrastive loss** is:

$$\mathcal{L}_{\mathrm{cl}} = -\frac{1}{2N}\left(\sum_{i=1}^{N}\log D_\psi(\boldsymbol{h}_i,\boldsymbol{s}) + \sum_{j=1}^{N}\log\left(1 - D_\psi(\boldsymbol{h}_j',\boldsymbol{s})\right)\right)$$

Surrogate EBM Training

We perform a surrogate optimization for our graph EBM in latent space:

$$\mathcal{L}_{\mathrm{ebm}} = \mathbb{E}_{\boldsymbol{h}\sim p_d(\boldsymbol{h})}[f_\omega(\boldsymbol{h})] - \mathbb{E}_{\tilde{\boldsymbol{h}}\sim p_\theta(\tilde{\boldsymbol{h}})}[f_\omega(\tilde{\boldsymbol{h}})] = \frac{1}{N}\sum_{i=1}^{N} f_\omega(\boldsymbol{h}_i) - \frac{1}{M}\sum_{j=1}^{M} f_\omega(\tilde{\boldsymbol{h}}_j)$$

Sampling from $p_\theta(h)$ can be achieved by ***efficient*** K-step MCMC sampling in latent space:

$$\tilde{\boldsymbol{h}}^{(k)} = \tilde{\boldsymbol{h}}^{(k-1)} - \lambda\nabla_{\tilde{\boldsymbol{h}}^{(k-1)}} f_\omega(\tilde{\boldsymbol{h}}^{(k-1)}) + \epsilon^{(k)}, \qquad \epsilon^{(k)} \sim \mathcal{N}(0,\sigma^2)$$

Multi-hop (MH) Graph Feature Encoder

To fully extract informative node representations and handle **heterophilic graphs**, we design a L-layer graph encoder:

$$\mathbf{X}^{(\ell)} = (\beta\mathbf{I} + \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2})\mathbf{X}^{(\ell-1)}, \qquad \mathbf{X}^{(0)} = \mathbf{X}$$

$$\mathbf{H} = g_\alpha(\mathbf{X},\mathbf{A}) = [\mathbf{H}^{(0)}||\mathbf{H}^{(1)}||\cdots||\mathbf{H}^{(L)}]\mathbf{W}_{\mathrm{enc}} + \boldsymbol{b}_{\mathrm{enc}}, \qquad \mathbf{H}^{(\ell)} = \mathbf{X}^{(\ell)}\mathbf{W}^{(\ell)} + \mathbf{b}^{(\ell)}$$

Contrastive Energy (CE)

Here we use readout $\boldsymbol{s}$ as supervised global prior information in energy head:

$$f_\omega(\boldsymbol{h}_i,\boldsymbol{s}) = \mathbf{W}[\boldsymbol{h}_i||\rho\boldsymbol{s}] + \boldsymbol{b}, \text{ or}$$
$$f_\omega(\boldsymbol{h}_i,\boldsymbol{s}) = \boldsymbol{h}_i^{\top}\mathbf{W}\boldsymbol{s}$$

Energy Readout (ERo)

To consider the importance of each node when computing readout, we use node energy to calculate weight $\bar{p} = \mathrm{softmax}(-f_\omega(\boldsymbol{h}_i))$, then

$$\boldsymbol{s}_p = \gamma\sum_i^N \bar{p}\boldsymbol{h}_i + (1-\gamma)\frac{1}{N}\sum_i^N \boldsymbol{h}_i$$

Recurrent Update

We design a **Recurrent Update** to promote ERo $\boldsymbol{s}_p$ and CE $f_\omega(\boldsymbol{h}_i,\boldsymbol{s})$ recurrently (see paper).

Experiments

Method	Cora Structure AUC↑	Cora Structure Acc↑	Cora Feature AUC↑	Cora Feature Acc↑	Cora Label AUC↑	Cora Label Acc↑	Amazon-Photo Structure AUC↑	Amazon-Photo Structure Acc↑	Amazon-Photo Feature AUC↑	Amazon-Photo Feature Acc↑	Amazon-Photo Label AUC↑	Amazon-Photo Label Acc↑
MSP	75.05	77.30	85.87	77.30	91.10	88.29	98.94	93.64	97.02	92.89	95.49	95.84
ODIN	30.57	74.60	21.19	77.50	20.29	87.66	3.50	91.42	5.31	92.91	10.16	96.08
Mahalanobis	41.03	71.90	63.92	74.20	67.45	88.92	62.40	93.42	72.47	92.88	60.80	95.84
Energy	79.48	79.10	89.34	79.30	93.26	90.19	99.94	93.04	98.51	92.76	97.13	95.68
ResidualFlow	61.23	82.20	60.72	82.20	64.69	82.20	82.58	93.55	75.46	93.55	78.31	93.55
GKDE	84.49	78.00	90.75	81.70	94.59	91.77	92.68	89.69	54.52	31.90	76.35	85.55
GPN	82.21	81.00	88.06	78.80	91.74	91.77	90.35	82.63	86.47	63.38	89.90	89.30
OODGAT	53.75	34.90	57.03	14.50	95.57	89.56	71.41	25.42	70.93	25.21	99.18	**96.16**
GNNSafe	87.98	75.30	92.18	75.40	92.36	88.92	98.69	93.74	98.47	92.96	97.34	95.72
DeGEM (Ours)	**99.93**	**84.20**	**99.84**	**84.30**	**97.58**	**93.04**	**100.00**	**94.49**	**99.91**	**93.97**	**99.28**	95.80

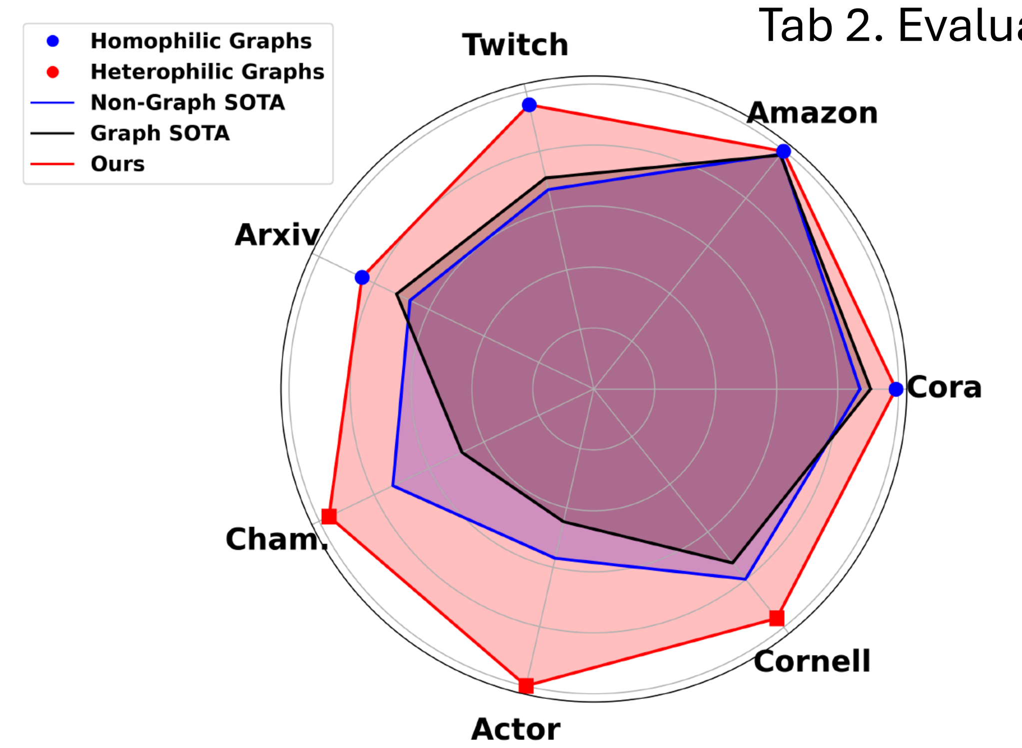
Method	Twitch ES AUC↑	Twitch FR AUC↑	Twitch RU AUC↑	Twitch Acc↑	Arxiv 2018 AUC↑	Arxiv 2019 AUC↑	Arxiv 2020 AUC↑	Arxiv Acc↑	Avg AUC↑	Avg Acc↑
MSP	66.08	47.40	57.41	63.89	61.76	63.52	67.06	53.87	83.07	80.38
ODIN	43.97	51.84	49.53	62.05	44.01	42.63	38.96	49.29	22.67	78.94
Mahalanobis	46.73	49.69	38.38	62.95	57.08	56.76	56.92	51.59	58.74	78.96
Energy	58.42	72.91	69.90	65.59	64.61	65.90	70.37	53.92	86.46	81.20
ResidualFlow	62.60	55.59	67.68	**66.37**	62.38	63.82	62.21	54.84	68.47	81.06
GKDE	57.52	57.48	46.25	60.42	69.44	71.32	71.73	21.15	77.25	67.52
GPN	84.07	76.32	78.36	60.29	OOM	OOM	OOM	OOM	N/A	N/A
OODGAT	77.35	77.72	73.24	60.29	72.35	73.97	72.30	54.69	74.61	50.09
GNNSafe	51.00	79.08	82.93	66.18	67.27	69.20	79.02	54.26	88.73	80.31
DeGEM (Ours)	**94.83**	**97.36**	**94.76**	64.51	**81.30**	**86.00**	**86.01**	58.20	**97.08**	**83.56**

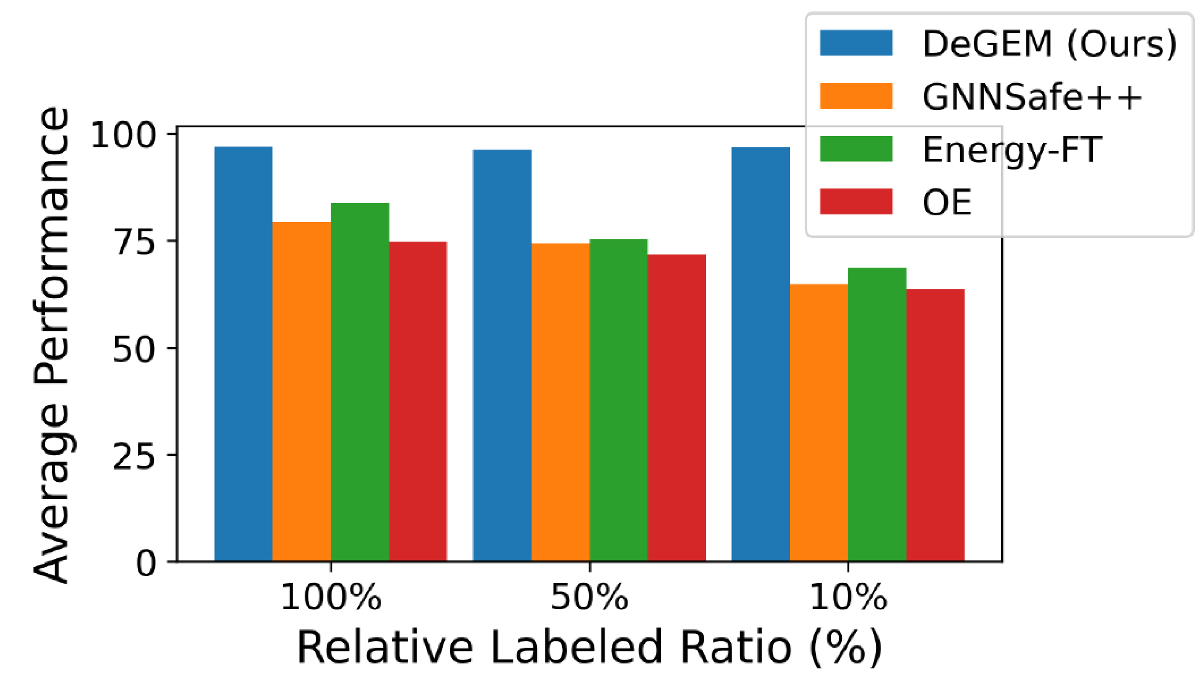
Tab 1. Evaluation on **homophilic** graphs

Method	Chameleon Structure AUC↑	Chameleon Structure Acc↑	Chameleon Feature AUC↑	Chameleon Feature Acc↑	Chameleon Label AUC↑	Chameleon Label Acc↑	Actor Structure AUC↑	Actor Structure Acc↑	Actor Feature AUC↑	Actor Feature Acc↑	Actor Label AUC↑	Actor Label Acc↑
MSP	99.28	33.41	70.92	32.42	53.32	43.89	71.38	21.64	59.11	24.34	56.32	37.27
ODIN	60.09	28.80	53.20	33.24	68.07	37.01	42.32	22.30	52.57	23.90	65.10	35.79
Mahalanobis	99.59	29.90	58.93	33.41	42.42	44.54	79.96	24.23	65.94	23.72	52.58	31.98
Energy	91.94	37.14	67.75	37.58	59.92	41.69	64.26	25.21	51.71	24.33	55.00	36.71
ResidualFlow	48.24	34.50	53.82	34.50	56.16	34.50	49.72	27.63	50.07	27.63	50.68	27.63
GKDE	96.06	30.12	67.33	34.94	60.22	40.40	71.27	25.63	58.08	19.56	53.72	33.60
GPN	82.90	20.41	64.99	30.99	72.68	34.71	78.58	18.67	62.13	20.21	75.04	**38.08**
OODGAT	54.89	26.06	53.86	29.73	65.33	40.68	51.25	23.65	52.00	25.77	65.39	36.30
GNNSafe	34.36	35.33	57.46	38.07	52.18	43.43	31.76	26.30	50.66	26.20	51.60	37.92
DeGEM (Ours)	**99.99**	**57.82**	**99.70**	**57.93**	**89.68**	**64.46**	**99.76**	**31.97**	**99.98**	**33.87**	**100.00**	36.02

Method	Cornell Structure AUC↑	Cornell Structure Acc↑	Cornell Feature AUC↑	Cornell Feature Acc↑	Cornell Label AUC↑	Cornell Label Acc↑	Avg AUC↑	Avg Acc↑
MSP	81.50	41.50	71.89	43.54	68.50	63.81	70.25	37.98
ODIN	67.99	38.78	85.76	38.10	42.16	63.81	59.69	35.75
Mahalanobis	69.84	36.73	70.08	43.54	81.73	62.86	69.01	36.77
Energy	83.09	38.10	86.70	38.10	69.70	62.86	70.01	37.97
ResidualFlow	67.80	43.54	66.33	43.54	73.09	43.54	57.32	35.22
GKDE	80.53	14.97	77.48	42.86	81.18	63.81	71.76	33.99
GPN	89.68	43.54	83.36	42.86	82.93	63.81	76.92	34.81
OODGAT	67.16	42.18	69.42	42.86	68.52	64.76	60.87	36.89
GNNSafe	74.66	25.17	76.22	41.50	68.17	63.81	55.23	37.52
DeGEM (Ours)	**97.97**	**48.30**	**100.00**	**65.31**	**90.90**	**77.14**	**97.55**	**52.53**

Tab 2. Evaluation on **heterophilic** graphs





Conclusion

We proposed **DeGEM** for node OOD detection task, overcoming the **heterophily issue** and **computational challenges** when MCMC sampling on large graphs. Extensive experiments demonstrate our superiority across datasets and under **few-label scenario**.

paper

code