Modeling Spectral Energy Shifts in Spatio-Temporal Graph Anomaly Detection

Yilin Liu[1], Hongchao Zhang[1], Taylor T. Johnson[1], Ahmad F. Taha[2], Meiyi Ma[1]

College of Connected Computing[1]
Department of Civil and Environmental Engineering[2]
Electrical and Computer Engineering[2]

yilin.liu.1@vanderbilt.edu

GitHub: https:/github.com/AICPS-Lab/Spectral-Energy-Shifts-in-GAD

EGNN Team

Yilin Liu

Hongchao Zhang

Taylor T. Johnson

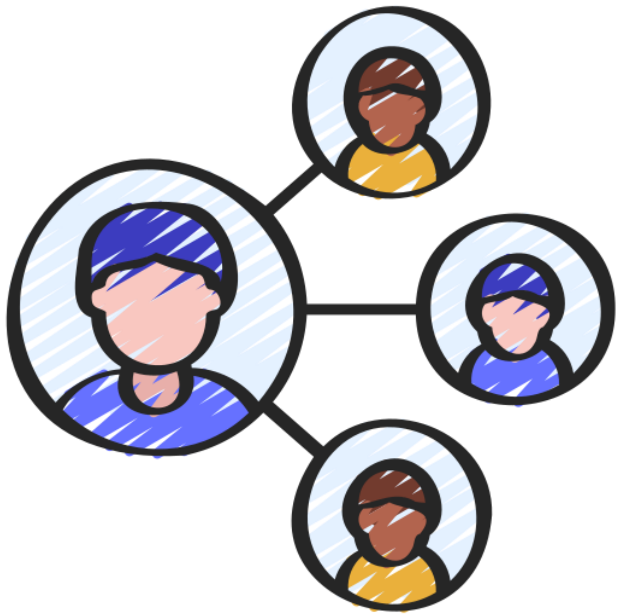
Ahmad F. Taha

Meiyi Ma

Motivation

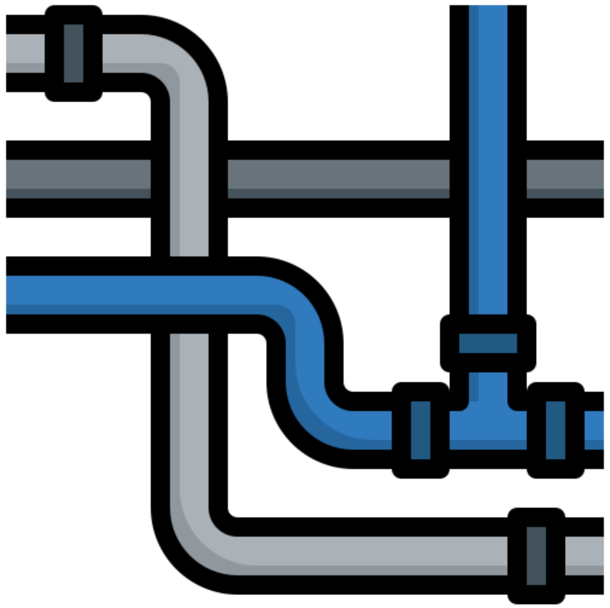
Graph Anomaly Detection

Graph Anomaly Detection (GAD): given a graph $G = (V, E)$ with node features $X \in \mathbb{R}^{N \times d}$,identify the set of anomalous nodes $V_a \subset V$ from normal nodes V_n.



Spam detection

Fraud detection



Leakage detection

Motivation

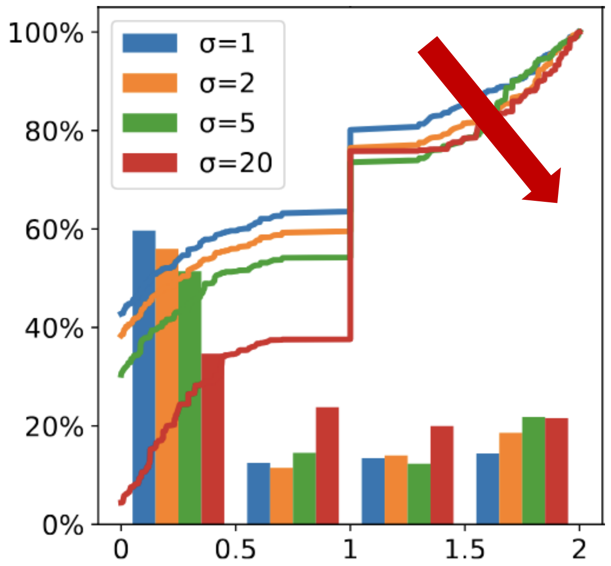
Recent Work

Recent studies[1] use the **spectral energy** to characterize anomalous behavior in the spectral domain.

$$\frac{x^T L x}{x^T x}$$

1. **Computation heavy**
2. **Not helpful in Message Passing NN**

x is 1-dim node feature; L is Laplacian matrix



Spectral energy of different degree of anomaly

Anomalies exhibit a **right-shift pattern.**

3. What about low variance anomaly (e.g. camouflage anomaly)?

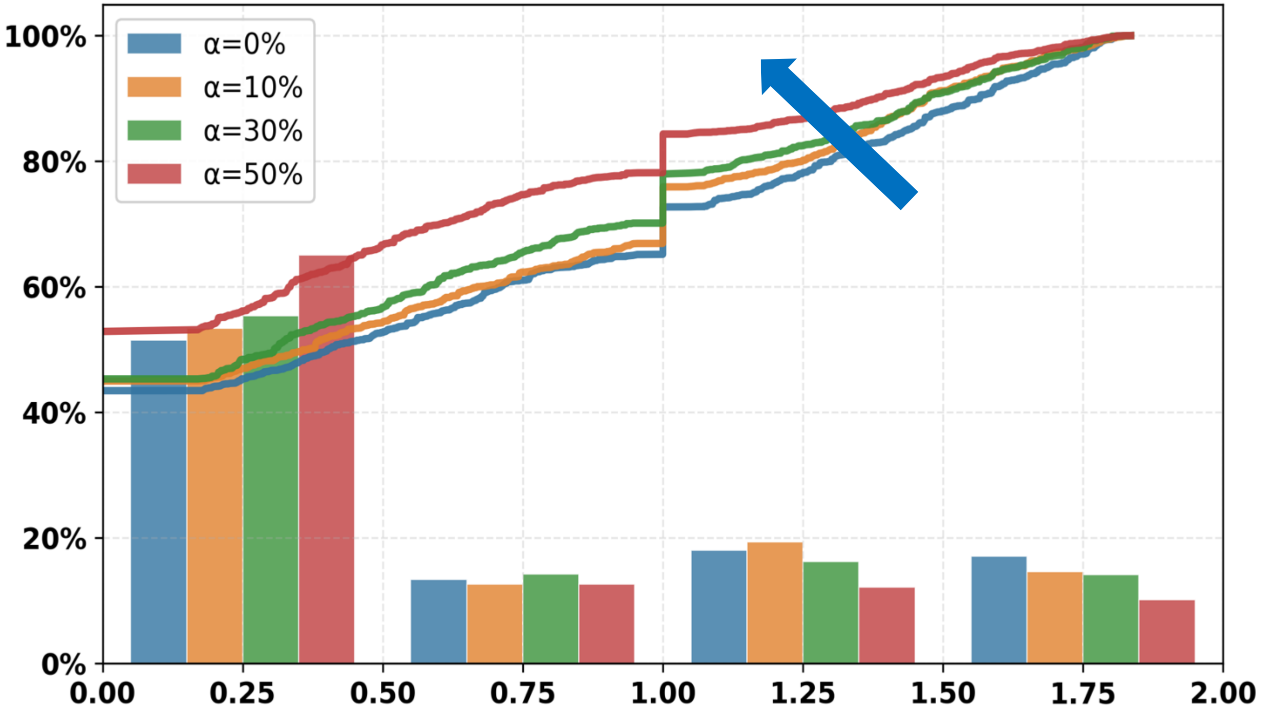
[1]Tang, Jianheng, et al. "Rethinking graph neural networks for anomaly detection." ICML 2022.

Contribution

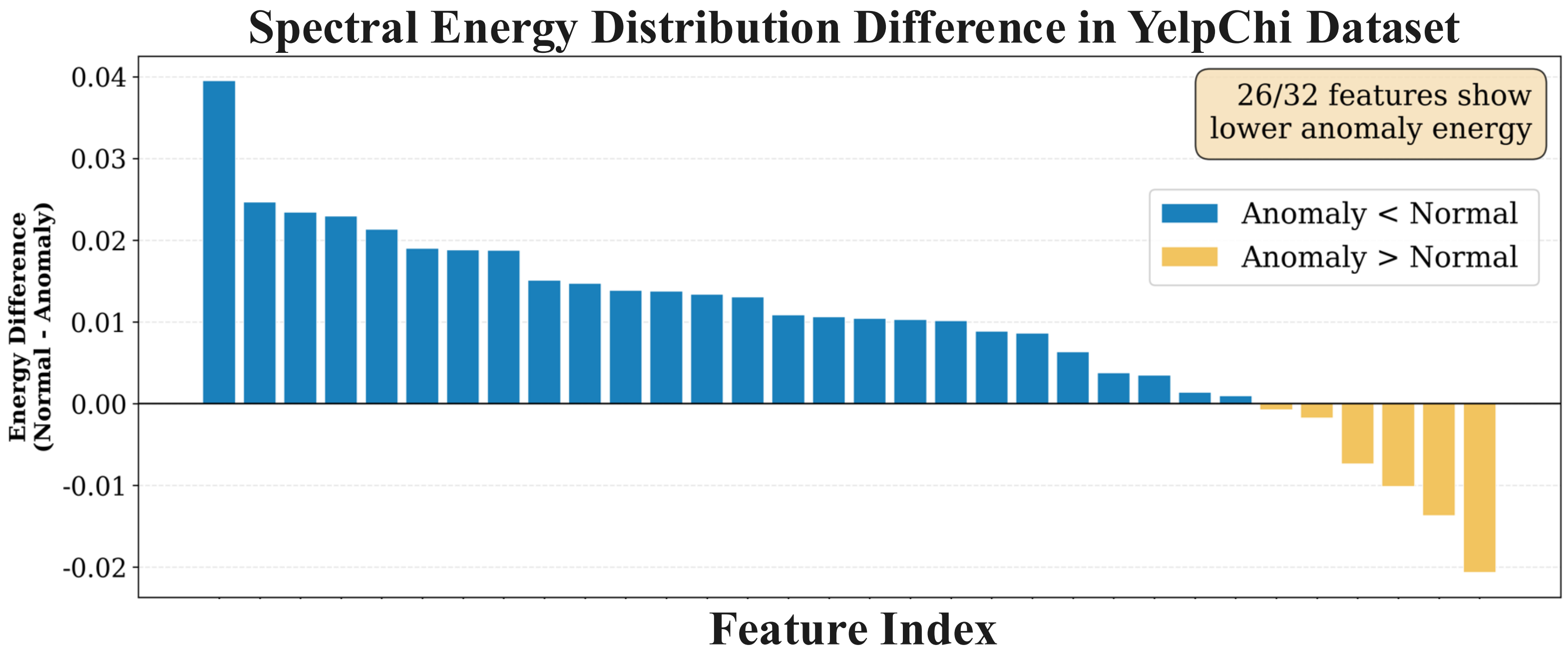
Our Contributions as follow:

- We identify a previously overlooked camouflaged anomaly pattern across multiple public datasets and characterize it through a spectral formulation.
- We propose EGNN, an energy-driven GAD method that able to capture the camouflaged anomalies scalable in large dataset.
- We generalize EGNN to time-series graphs robust in sliding windows and data-efficient.
- EGNN achieves competitive performance against SOTAs across multiple benchmarks.

Camouflage Anomalies

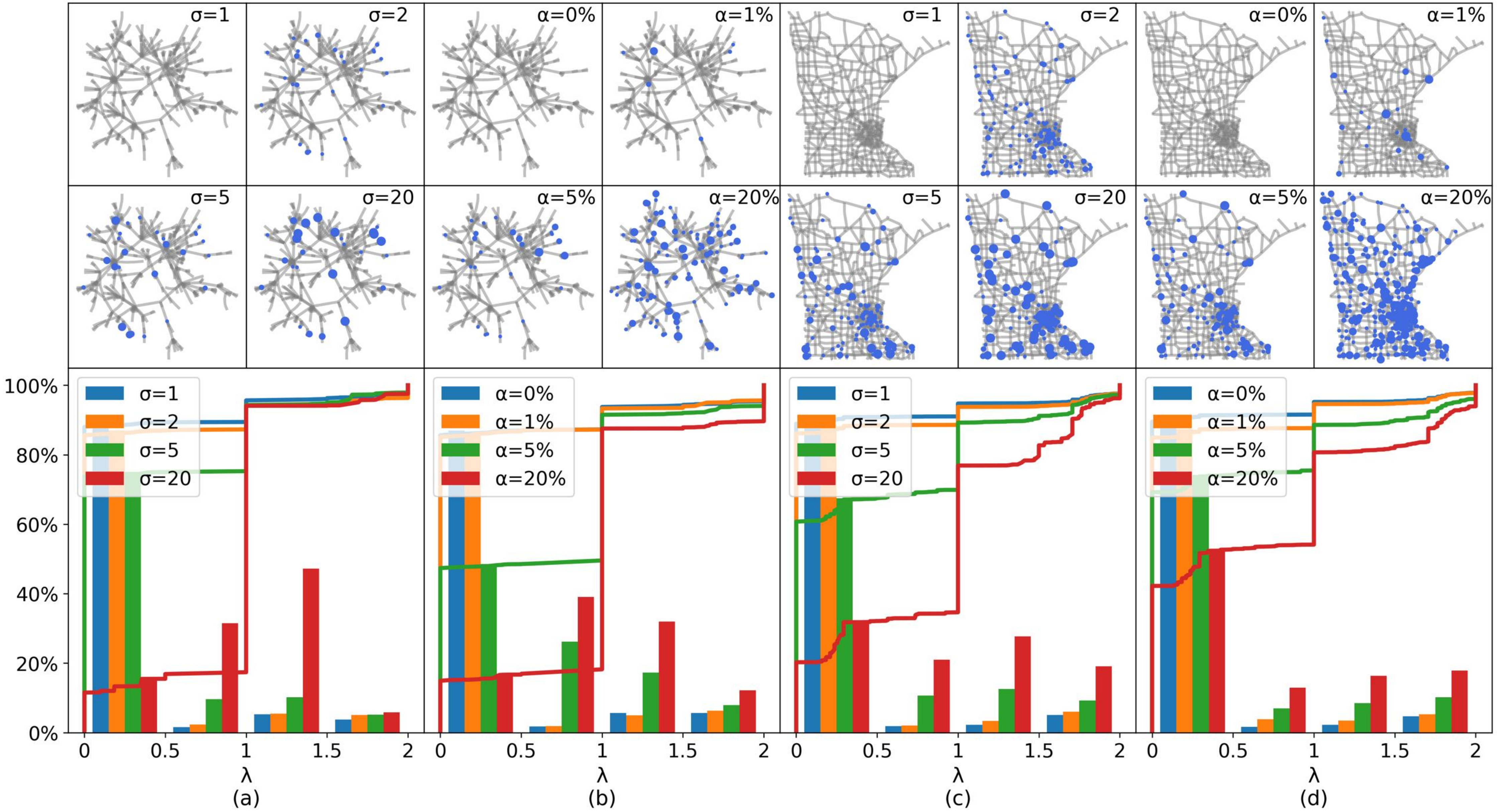


Camouflage anomalies exhibit a **left-shift pattern**



Camouflage anomalies **commonly exist** in real-world dataset

Local Approximation



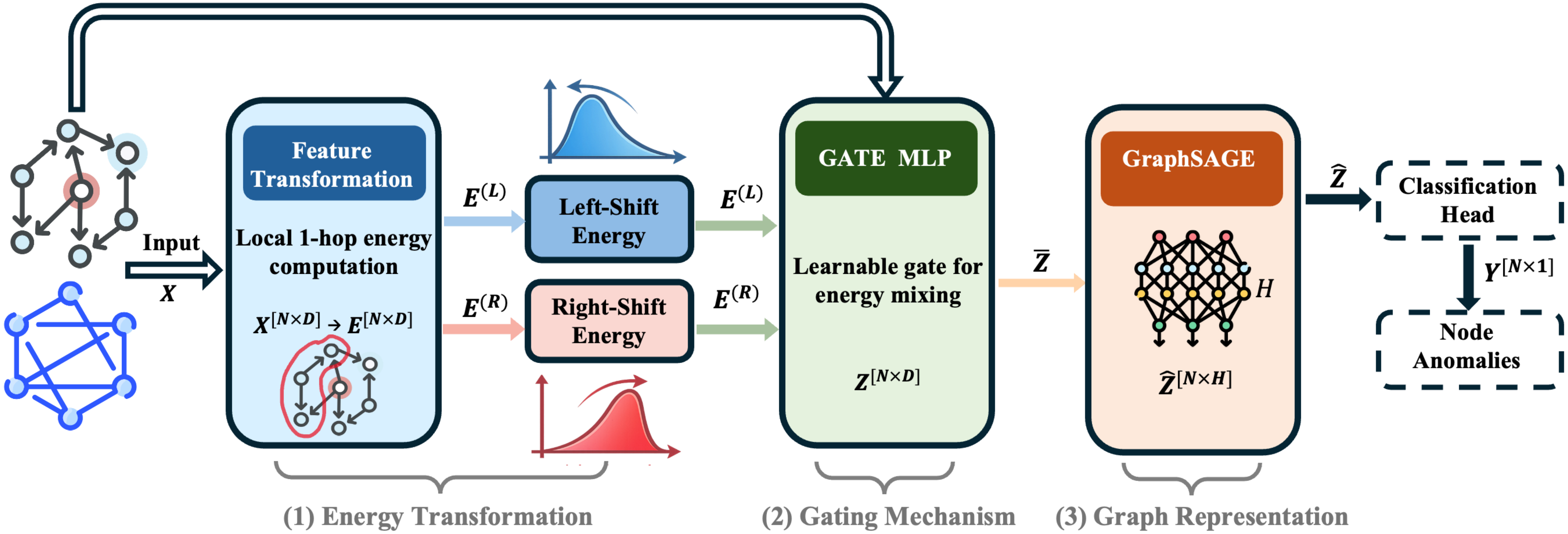
Local spectral anomalies under varying anomaly variance

Left-Shift Spectral Energy

$$E^{(R)}_{\mathcal{N}_i,f} = \frac{\mathbf{x}^\top_{\mathcal{N}_i,f} L_{\mathcal{N}_i} \mathbf{x}_{\mathcal{N}_i,f}}{\mathbf{x}^\top_{\mathcal{N}_i,f} \mathbf{x}_{\mathcal{N}_i,f}} \qquad E^{(L)}_{\mathcal{N}_i,f} = \frac{\mathbf{x}^\top_{\mathcal{N}_i,f} (2I - L_{\mathcal{N}_i}) \mathbf{x}_{\mathcal{N}_i,f}}{\mathbf{x}^\top_{\mathcal{N}_i,f} \mathbf{x}_{\mathcal{N}_i,f}} = 2 - E^{(R)}_{\mathcal{N}_i,f}$$

L_N *: denotes the (normalized) Laplacian of the induced subgraph.*
f *: feature*
X *: feature vector*

EGNN Framework



Evaluation in Static Graph

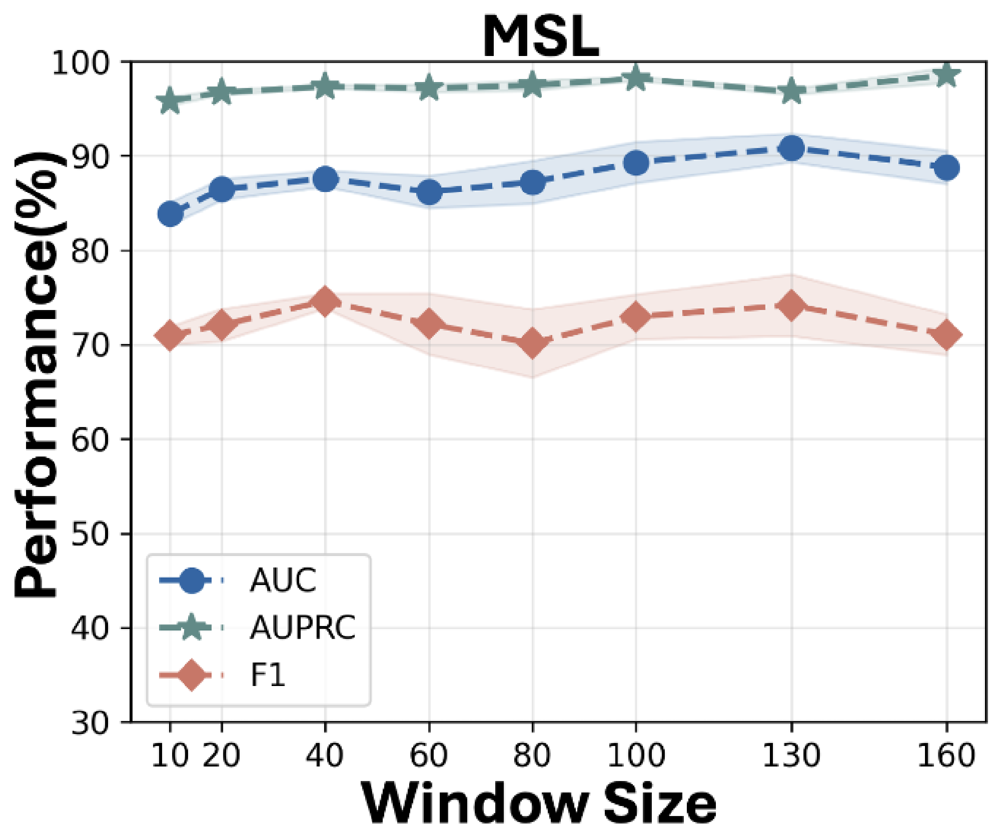
Model	Amazon			YelpChi			T-Finance			T-Social			
	F1-m	AUC	PRC	F1-m	AUC	PRC	F1-m	AUC	PRC	F1-m	AUC	PRC	Time(s)
GCN	63.22	82.20	34.15	55.70	58.91	22.42	72.63	85.59	43.45	65.34	78.80	23.96	2041
ChebyNet	91.03	95.76	86.18	71.59	84.30	54.46	85.70	93.29	75.37	58.31	73.56	12.04	1380
GAT	85.70	93.29	75.37	65.92	78.33	41.34	81.56	91.69	55.90	67.08	85.47	25.65	2945
GIN	88.93	92.32	80.29	65.20	76.59	39.51	79.65	84.41	51.83	52.61	66.37	5.92	3611
GraphSAGE	75.28	88.90	66.20	68.59	82.17	45.80	57.04	57.98	8.71	58.42	74.04	9.44	3063
SGC	63.06	77.10	23.68	51.35	51.91	15.76	57.92	51.91	15.76	42.49	48.93	3.04	2109
GT	89.14	90.75	76.20	67.27	79.86	44.89	64.63	78.50	19.87	63.66	83.43	20.82	2296
BernNet	91.31	93.83	84.01	69.19	82.14	48.89	81.48	90.66	52.90	51.94	64.68	4.88	1945
PC-GNN	67.04	81.80	29.36	56.30	59.55	22.91	85.62	92.17	73.70	51.19	72.88	13.77	15399
BWGNN	91.40	96.17	86.42	71.78	84.33	54.67	84.67	93.12	73.84	81.78	94.32	60.81	3045
UniGAD	90.46	**96.60**	86.65	71.23	83.69	53.97	89.34	95.14	**84.71**	78.67	91.65	58.33	3981
Ours	**91.52**	96.32	**89.40**	**76.89**	**88.10**	**66.26**	**89.60**	**95.39**	84.39	**95.40**	**99.69**	**95.89**	3712

T-social with 5,781,065 nodes and 73,105,508 edges.

EGNN improve F1-m more than **16%** on the hardest and largest T-Social dataset!

Evaluation in Time-Series Graph

Model	MSL			SWaT			WADI			Avg. F1	#Params
	AUC	PRC	F1-m	AUC	PRC	F1-m	AUC	PRC	F1-m		
TranAD	73.57	92.71	68.14	81.49	70.62	76.45	43.53	5.24	48.50	64.36	261,243
USAD	80.60	95.02	63.71	78.21	69.43	71.52	49.87	5.63	48.35	61.19	1.28M
GDN	65.79	88.80	57.56	80.57	**71.47**	75.45	46.99	4.96	48.18	60.39	5,121
EGNN	**86.63**	**97.28**	**70.50**	**93.70**	70.11	**86.03**	**67.88**	**20.27**	**61.41**	**72.65**	53,953



EGNN achieves the **highest performance** on time-series datasets and remains **robust in sliding-window**, show **data efficiency**.

Takeaway

- We **identify camouflaged anomalies** as a new pattern that induces a **left-shift** in spectral energy distributions.
- Building on this insight, **we propose EGNN**, an energy-aware GAD framework that expands the detection of node attribute anomalies and **extends efficiently to temporal graphs**.
- Experiments show that **EGNN consistently outperforms state-of-the-art methods** with orders of magnitude fewer parameters.

ICML
International Conference On Machine Learning

VANDERBILT School *of* Engineering

Thank You!