ICML | 2022

Thirty-ninth International Conference on Machine Learning

Rethinking Graph Neural Networks for Anomaly Detection

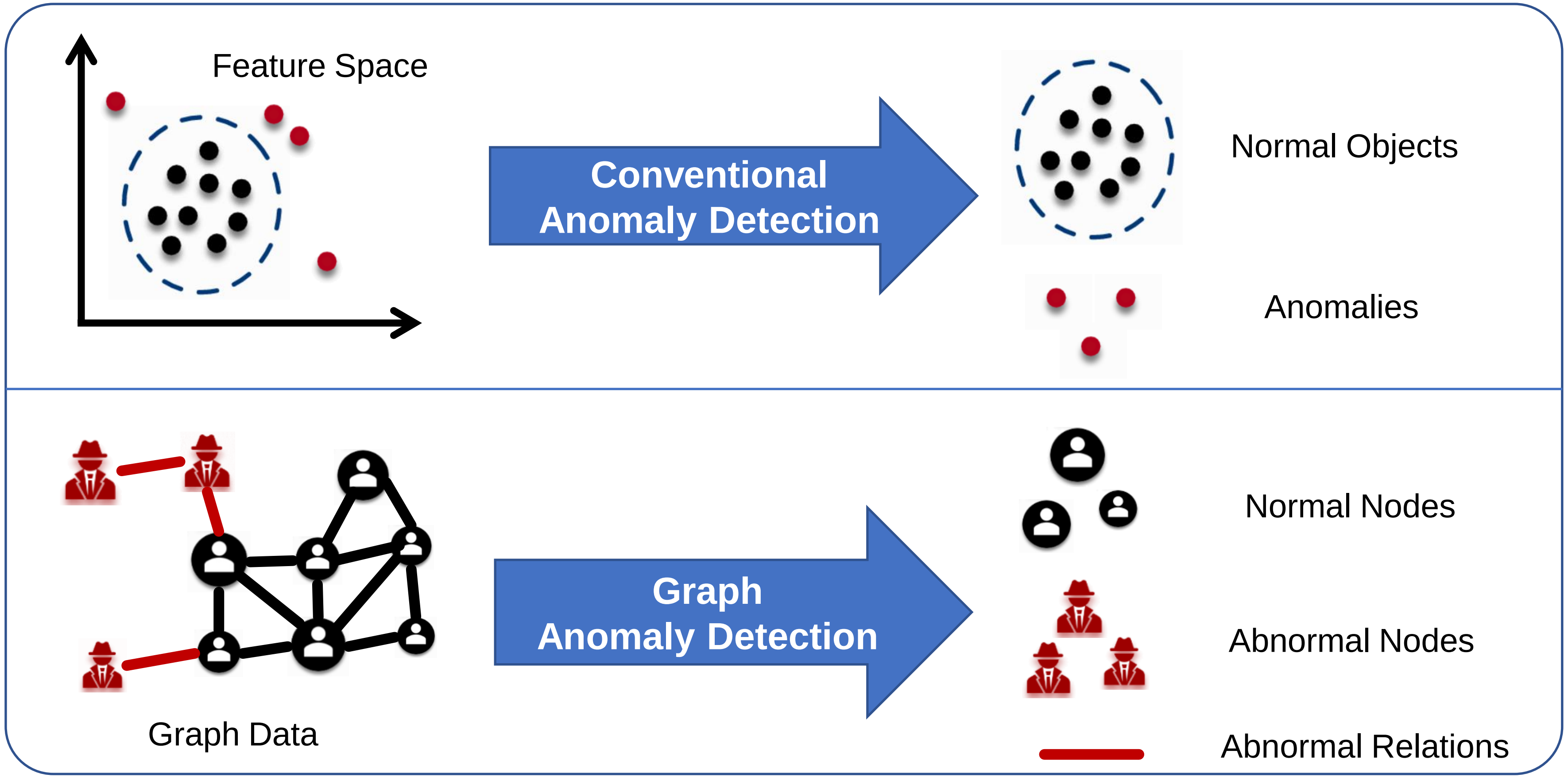
Jianheng Tang[1,2], Jiajin Li[3], Ziqi Gao[1,2], Jia Li[1,2]

[1]Hong Kong University of Science and Technology (Guangzhou), [2]Hong Kong University of Science and Technology, [3]Stanford University

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Task Introduction

Graph anomaly detection aims to identify **abnormal graph objects** that **deviate significantly from the majority**, e.g, identifying fraudsters in a transaction network. [1]



[1] Ma et al. A comprehensive survey on graph anomaly detection with deep learning. TKDE’21.

Motivation

- Graph Neural Networks (GNNs) are widely used for this task.
- Existing methods analyze the anomaly detection from the **graph spatial domain**. There are few works that address this problem from **the spectral domain**.
- The spectral filter determines the expressive power of GNN [1].

Spatial GNN $$H_{:v}^{(l+1)} = upd\Big(g_0(H_{:v}^{(l)}), agg\Big(g_1(H_{:u}^{(l)}) : u \in \mathcal{N}(v)\Big)\Big)$$

Spectral GNN $$g_\theta(\mathbf{L}) * x = \mathbf{U} g_\theta^*(\mathbf{\Lambda}) \mathbf{U}^\top \mathbf{x},$$

[1] Balcilar et al. Analyzing the expressive power of graph neural networks in a spectral perspective. ICLR’21

Motivation

- Graph Neural Networks (GNNs) are widely used for this task.
- Existing methods analyze the anomaly detection from the **graph spatial domain**. There are few works that address this problem from **the spectral domain**.
- The spectral filter determines the expressive power of GNN [1].

How to choose a tailored spectral filter in GNN for anomaly detection?

- We take the first step towards analyzing anomalies via the lens of the **graph spectrum** (i.e., after the graph Fourier transform of node attributes).

Key Observation

The existence of anomalies leads to the ‘**right-shift**’ of spectral energy, which means the spectral energy distribution **concentrates less in low frequencies and more in high frequencies**.

Theoretical Contribution

The existence of anomalies leads to the '**right-shift**' of spectral energy, which means the spectral energy distribution **concentrates less in low frequencies and more in high frequencies**.

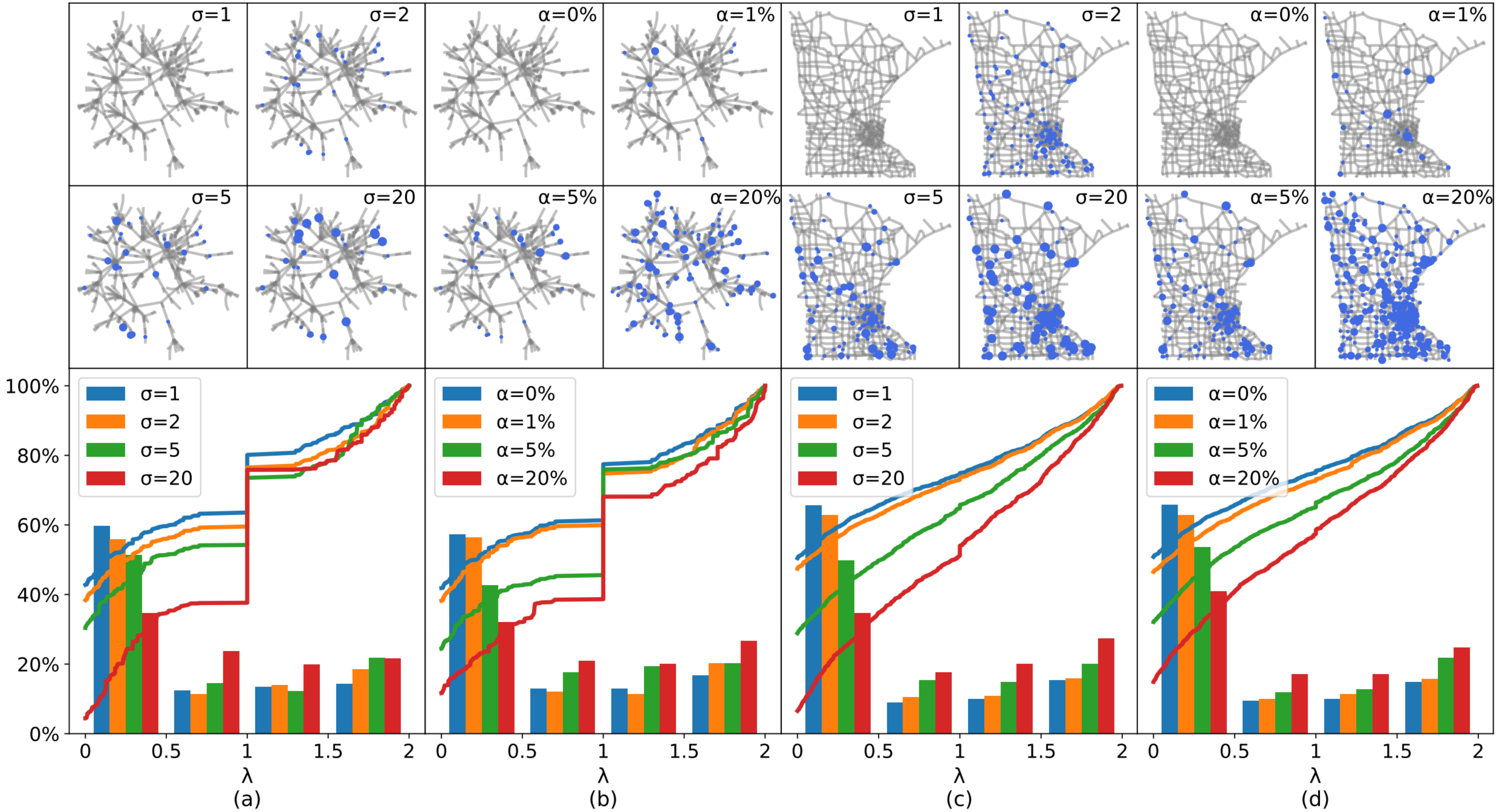
We prove it on a Gaussian anomaly model in a rigorous manner:

Proposition 2. *If* $|\mu| \neq 0$ *and* $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, *the expectation of the inverse of low-frequency energy ratio* $\mathbb{E}_{\boldsymbol{x}}[1/\eta_k(\boldsymbol{x}, \boldsymbol{L})]$ *is monotonically increasing with the anomaly degree* $\sigma/|\mu|$.

Spectral Analysis of Graph Anomaly

Anomalies in the graph spatial domain

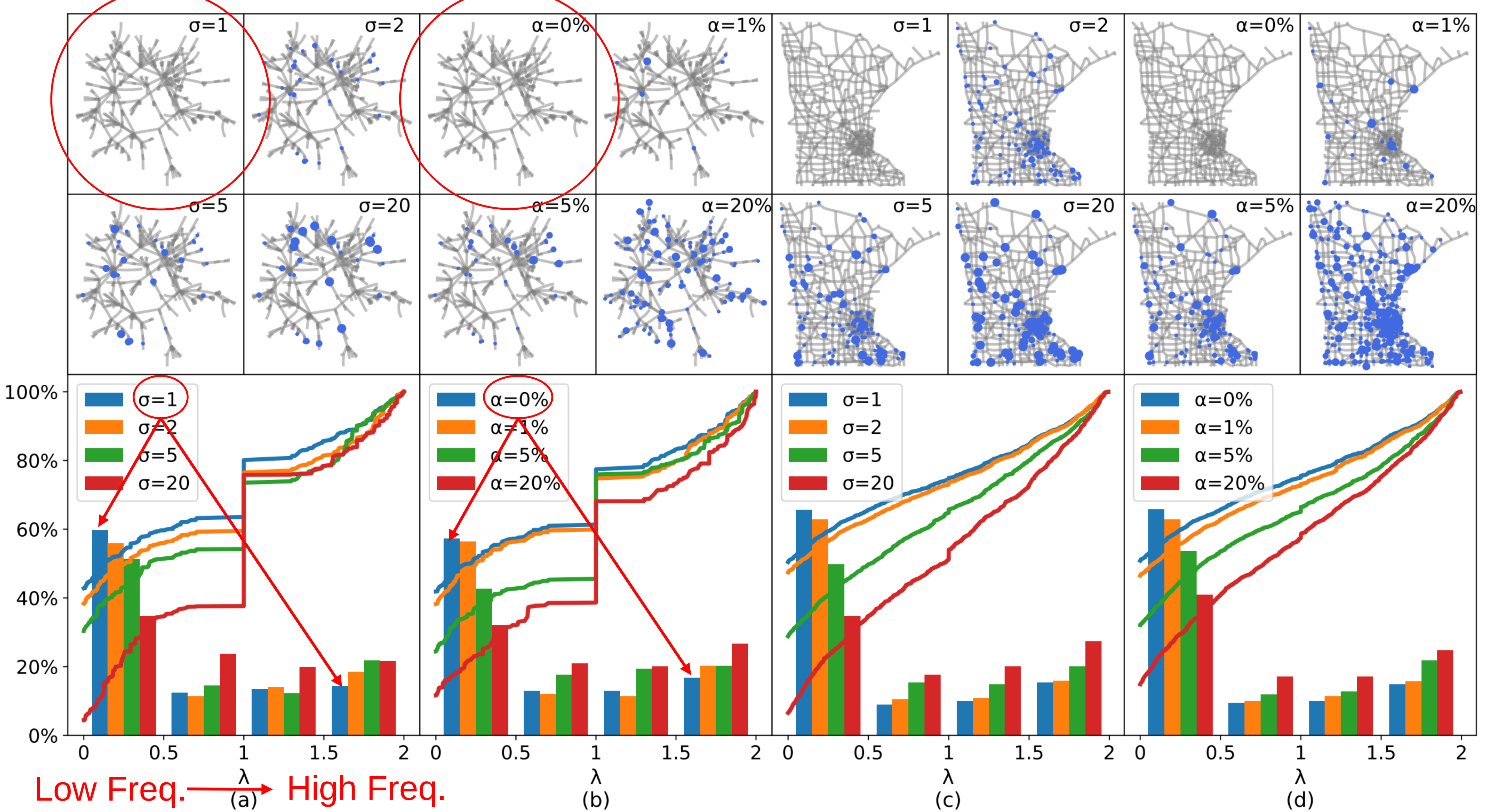
Corresponding spectral energy distribution



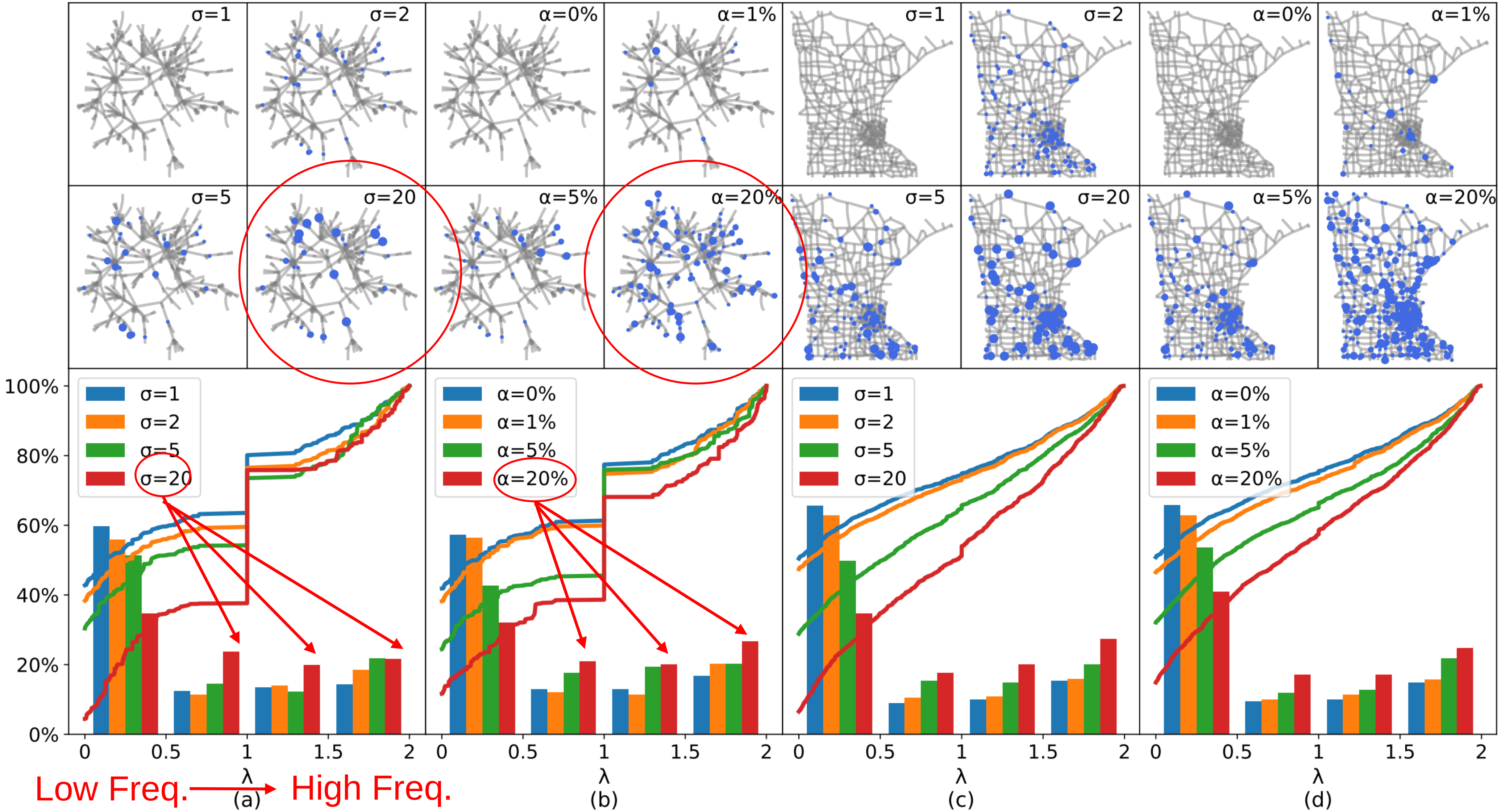
Barabasi Albert graph

Minnesota road graph

Spectral Analysis of Graph Anomaly



Spectral Analysis of Graph Anomaly



Validation on Real-world Anomalies

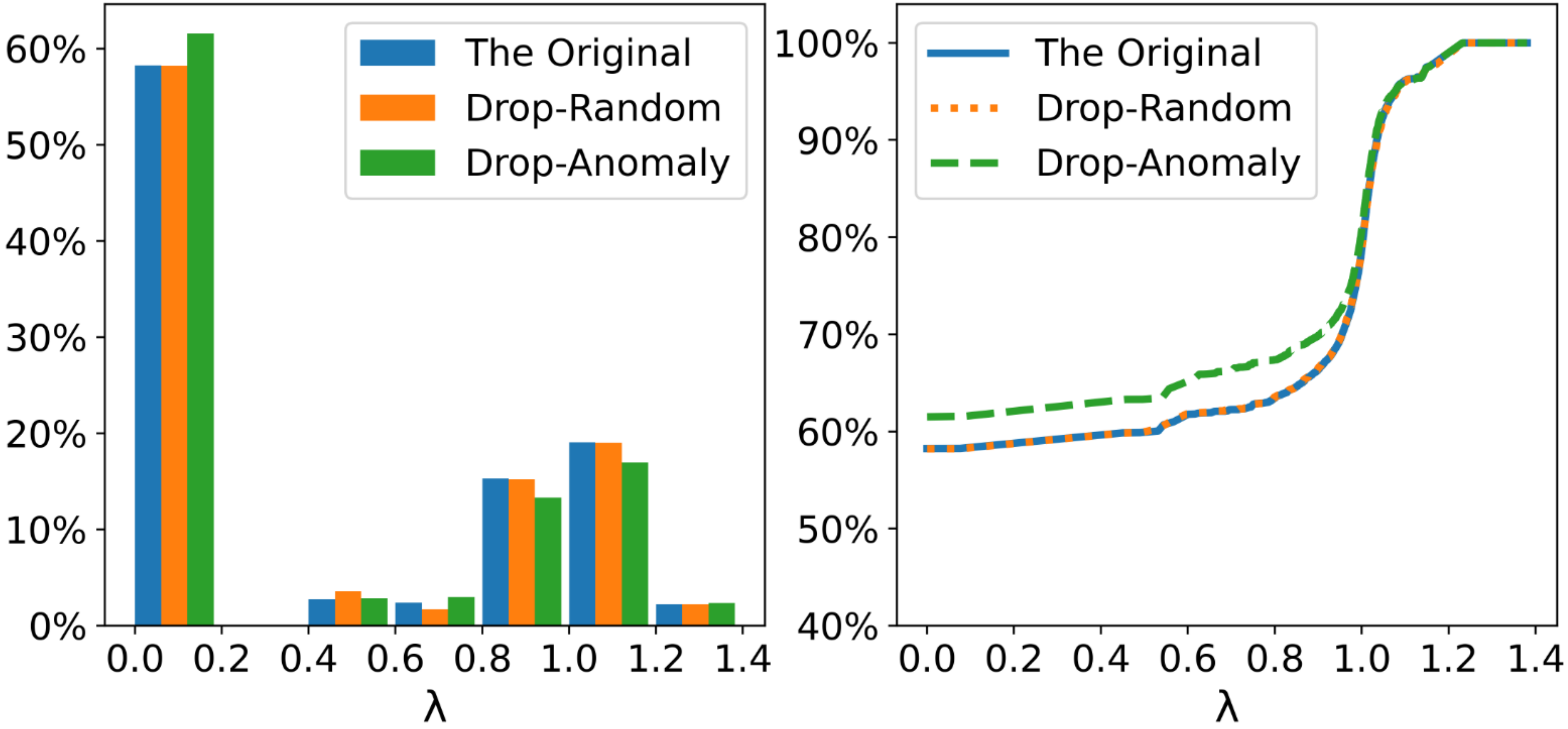


Figure 3. Comparison of the spectral energy distribution (left) and the energy ratio curve (right) between the original graph and two perturbed graphs in the Amazon dataset.

Methodology

The analysis of graph anomaly shows that we need to focus on ‘right-shift’ effect when detecting graph anomalies.

Unfortunately, most of the current GNNs are low-pass filters or adaptive filters that are neither guaranteed to be band-pass nor spectral-localized.

We propose our new graph neural network architecture based on **Hammond's graph wavelet theory**[1], which is band-pass in nature and can better address the `right-shift' effect inheriting from anomalies.

[1] Hammond et al. Wavelets on graphs via spectral graph theory. Applied and Computational Harmonic Analysis.

Beta Wavelet

A group of $C+1$ Beta wavelets with the same order:

$$\mathcal{W} = (\mathcal{W}_{0,C}, \mathcal{W}_{1,C-1}, \cdots, \mathcal{W}_{C,0})$$

$$\mathcal{W}_{p,q} = \boldsymbol{U}\beta^*_{p,q}(\boldsymbol{\Lambda})\boldsymbol{U}^T = \beta^*_{p,q}(\boldsymbol{L}) = \frac{(\frac{\boldsymbol{L}}{2})^p(I - \frac{\boldsymbol{L}}{2})^q}{2B(p+1, q+1)}$$

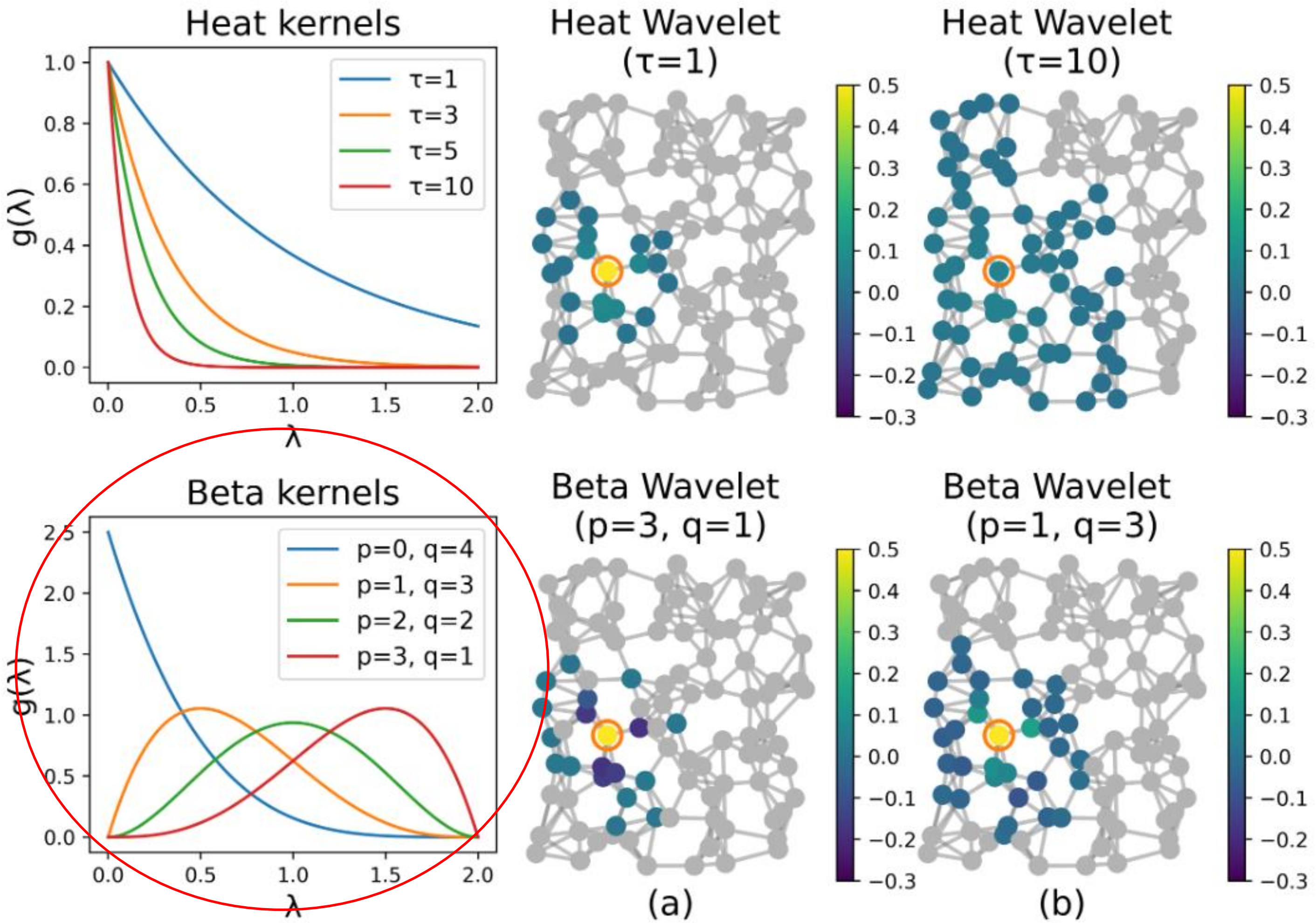
Derived From Beta distribution

$$\beta_{p,q}(w) = \begin{cases} \frac{1}{B(p+1,q+1)} w^p (1-w)^q & \text{if } w \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

Good Properties of Beta Kernel

Compared with the widely used Heat kernel [1], Beta kernel has the following good properties:

- Band-pass and spectral-localized filters

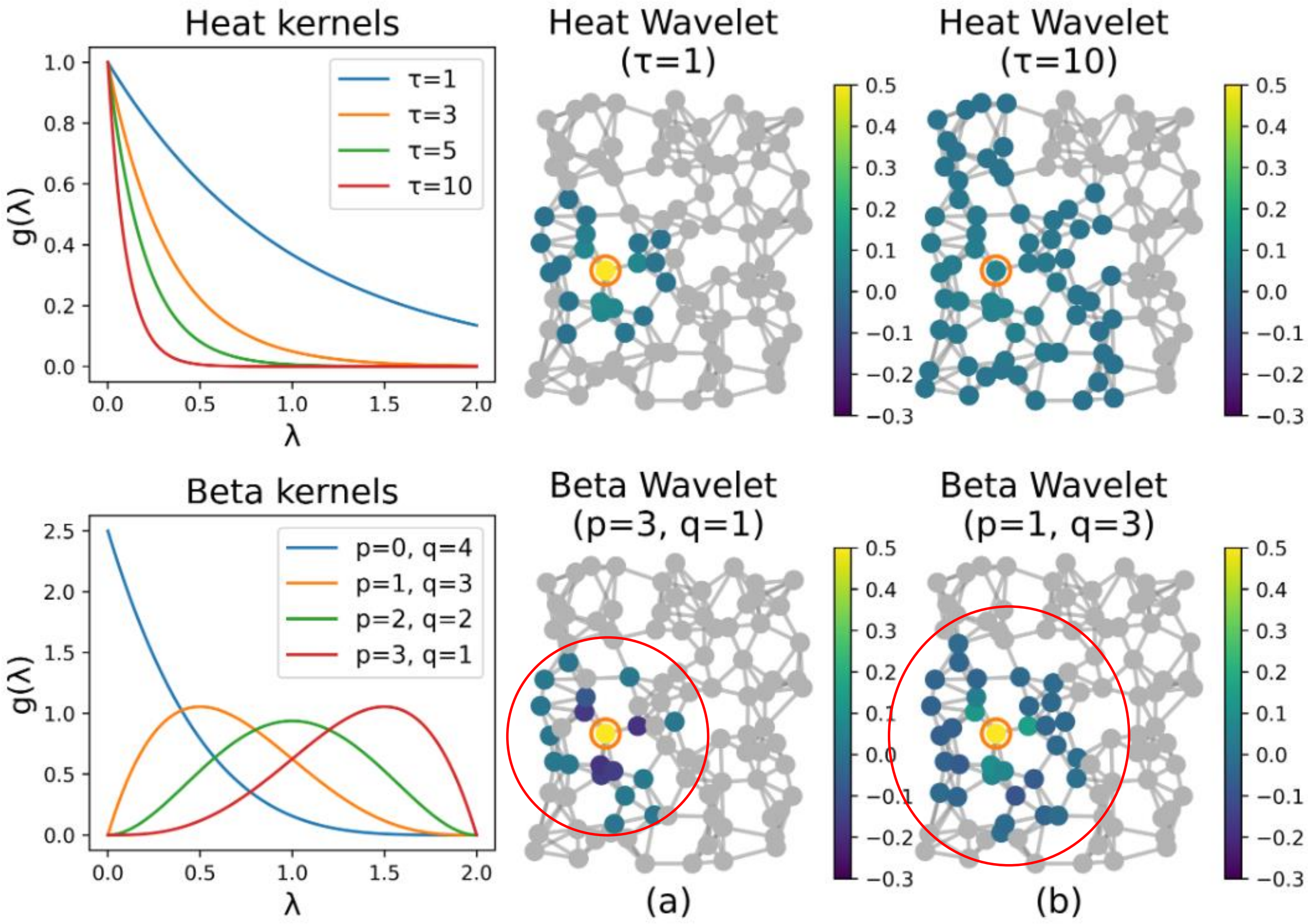


[1] Xu et al. Graph wavelet neural network. ICLR’19

Good Properties of Beta Kernel

Compared with the widely used Heat kernel [1], Beta kernel has the following good properties:

- Band-pass and spectral-localized filters
- Spatial locality
- Support fast computation



[1] Xu et al. Graph wavelet neural network. ICLR'19

Beta Wavelet Graph Neural Network (BWGNN)

A group of $C+1$ Beta wavelets with the same order:

$$\mathcal{W} = (\mathcal{W}_{0,C}, \mathcal{W}_{1,C-1}, \cdots, \mathcal{W}_{C,0})$$

$$\mathcal{W}_{p,q} = \boldsymbol{U}\beta^*_{p,q}(\boldsymbol{\Lambda})\boldsymbol{U}^T = \beta^*_{p,q}(\boldsymbol{L}) = \frac{(\frac{\boldsymbol{L}}{2})^p(I-\frac{\boldsymbol{L}}{2})^q}{2B(p+1,q+1)}$$

Beta Wavelet Graph Neural Network:

$$Z_i = \mathcal{W}_{i,C-i}(\mathrm{MLP}(\boldsymbol{X})) \longrightarrow$$

Feed input $\boldsymbol{X}$ to different wavelets in parallel

$$\boldsymbol{H} = \mathrm{AGG}([Z_0, Z_1, \cdots, Z_C]) \longrightarrow$$

Aggregate all filtering results

Datasets

Dataset	Statistics			
	# Nodes	# Edges	Anomaly(%)	# Features
Amazon	11,944	4,398,392	6.87%	25
YelpChi	45,954	3,846,979	14.53%	32
T-Finance	39,357	21,222,543	4.58%	10
T-Social	5,781,065	73,105,508	3.01%	10

Table 1. Summary of dataset statistics

Results on YelpChi and Amazon

Table 4. Experimental results (Mean ± Std.) of compared methods on the YelpChi and Amazon datasets with 1% and 40% training ratios.

Dataset	YelpChi (1%)		YelpChi (40%)		Amazon (1%)		Amazon (40%)	
Metric	F1-macro	AUC	F1-macro	AUC	F1-macro	AUC	F1-macro	AUC
MLP	53.90±0.23	59.83±0.40	57.57±0.89	66.52±1.09	74.68±1.25	83.62±1.76	79.17±1.26	89.80±1.04
SVM	60.47±0.24	62.92±0.92	70.77±0.01	70.37±0.04	83.49±1.39	81.62±3.53	90.71±0.04	90.51±0.07
GCN	52.48±0.50	54.06±0.72	54.31±0.77	56.51±1.09	67.93±1.42	82.85±0.71	67.47±0.52	83.49±0.47
ChebyNet	63.13±0.50	73.48±0.74	65.72±0.48	78.19±0.63	85.74±1.67	87.60±0.61	91.94±0.29	94.64±0.53
GAT	50.27±2.31	50.95±1.39	54.64±2.19	57.20±0.24	60.84±2.47	73.45±1.26	83.18±2.91	89.90±0.95
GIN	57.57±1.15	64.73±1.73	62.85±0.76	74.09±1.06	68.69±4.12	78.83±3.82	69.26±2.45	80.56±2.99
GraphSAGE	58.41±2.12	67.58±1.69	65.49±1.84	78.31±2.14	70.78±3.85	75.37±2.49	74.17±1.37	86.95±2.74
GWNN	59.10±6.53	67.16±11.44	65.29±6.67	75.32±8.97	87.01±1.98	85.37±2.32	91.00±0.27	93.19±2.22
GraphConsis	56.79±2.72	66.41±3.41	58.70±2.00	69.83±3.02	68.59±3.41	74.11±3.53	75.12±3.25	87.41±3.34
CAREGNN	62.18±1.39	75.07±3.88	63.32±0.94	76.19±2.92	68.78±1.68	88.69±3.58	86.39±1.73	90.53±1.67
PC-GNN	59.82±1.42	75.47±0.98	63.00±2.30	79.87±0.14	79.86±5.65	**90.40±2.05**	89.56±0.77	95.86±0.14
BWGNN (Homo)	61.15±0.41	72.01±0.48	71.00±0.91	84.03±0.98	**90.92±0.78**	89.45±0.33	**92.29±0.44**	**98.06±0.45**
BWGNN (Hetero)	**67.02±0.50**	**76.95±1.38**	**76.96±0.89**	**90.54±0.49**	83.83±3.79	86.59±2.62	91.72±0.84	97.42±0.48

Results on T-Finance and T-Social

Table 3. Experimental results and the overall training time (seconds) on the T-Finance and T-Social datasets with different training ratios.

Dataset	T-Finance (1%)		T-Finance (40%)			T-Social (0.01%)		T-Social (40%)		
Metric	F1-macro	AUC	F1-macro	AUC	Time	F1-macro	AUC	F1-macro	AUC	Time
MLP	61.00	82.93	70.57	87.15	13.32	50.03	56.35	50.35	56.96	986
SVM	67.69	71.47	76.23	78.16	145.11	57.69	50.06	-	-	>1 day
GCN	54.11	57.30	70.74	64.43	23.98	49.23	59.04	59.88	87.35	1294
ChebyNet	77.20	85.53	80.81	88.45	26.13	52.59	70.02	64.77	85.52	1711
GAT	53.15	52.04	53.86	73.00	181.62	46.25	44.35	69.01	89.06	1596
GIN	58.25	68.86	65.23	80.02	32.39	58.32	70.61	61.74	79.72	2195
GraphSAGE	59.03	66.35	52.71	67.12	35.91	57.91	59.69	59.77	70.80	2230
GWNN	70.64	86.68	71.58	86.57	27.25	50.81	56.14	58.72	73.77	1992
GraphConsis	71.73	90.28	73.46	91.42	264.41	52.45	65.29	56.55	71.25	3495
CAREGNN	73.32	90.50	77.55	92.16	572.41	55.82	71.20	56.26	71.86	9159
PC-GNN	62.06	90.76	63.18	91.23	736.55	51.14	59.84	52.17	68.45	13958
BWGNN	**84.89**	**91.15**	**86.87**	**94.35**	31.98	**75.93**	**88.06**	**83.98**	**95.20**	2707

Conclusion

- We presents a novel analysis of graph anomalies in the spectral domain.
- We find that graph anomalies lead to the ‘right-shift’ phenomenon of spectral energy distributions
- We propose Beta Wavelet Graph Neural network (BWGNN) to better capture anomaly information on graph.
- Our code and data are available at
  https://github.com/squareRoot3/Rethinking-Anomaly-Detection
- Feel free to contact me via email (jtangbf@connect.ust.hk) or GitHub.