SPECFORMER: SPECTRAL GRAPH NEURAL NETWORKS MEET TRANSFORMERS

Deyu Bo[1], **Chuan Shi**[1*], **Lele Wang**[2], **Renjie Liao**[2*]
Beijing University of Posts and Telecommunications[1],
University of British Columbia[2]
{bodeyu, shichuan}@bupt.edu.cn, {lelewang, rjliao}@ece.ubc.ca

1. Introduction and Contributions –

- GNNs are either *spatial* or *spectral.*
- Spatial GNNs are heavily explored.
- (Goal) Build expressive spectral filters that counteract both current problems.

- (Contribution) Novel **set-to-set** spectral filter generator.
- (Contribution) generated filters have **good properties**.
- (Contribution) SOTA performance on synthetic datasets, based on filter recovery.
- (Contribution) near-SOTA on most datasets.

2. Related Work –

- **Spectral GNNs** with limitations on capturing relative dependencies
- **Graph Transformers** that are completely spatial
- $\tilde{x} = \mathrm{U}G_\theta U^T x$ task is to make the filter powerful enough to capture relative dependencies.

3. Specformer –

- Eigenvalue Encoding

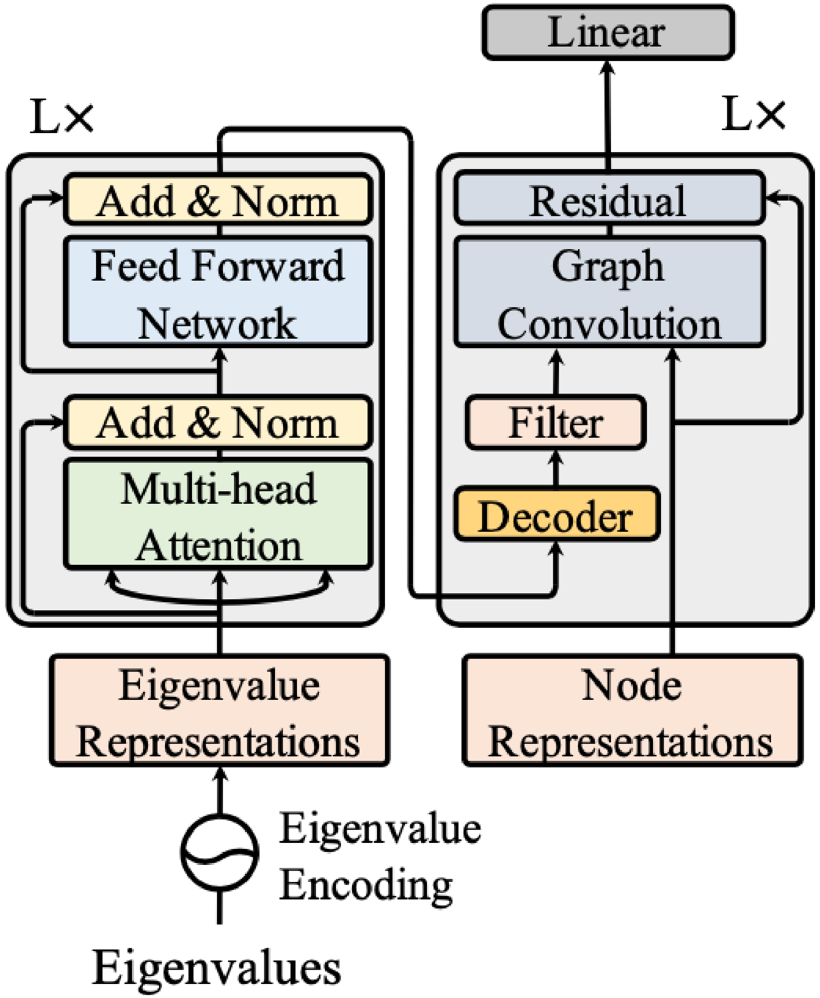
  1. Overcome scalar-inexpressivity problem

$$\rho(\lambda, 2i) = \sin(\epsilon\lambda/10000^{2i/d})$$
$$\rho(\lambda, 2i+1) = \cos(\epsilon\lambda/10000^{2i/d})$$

  2. Normal bi-directional Transformer on top

$$\boldsymbol{Z} = [\lambda_1 \| \rho(\lambda_1), \cdots, \lambda_n \| \rho(\lambda_n)]^\top \in \mathbb{R}^{n \times (d+1)}$$

$$\tilde{\boldsymbol{Z}} = \mathrm{MHA}(\mathrm{LN}(\boldsymbol{Z})) + \boldsymbol{Z},$$
$$\hat{\boldsymbol{Z}} = \mathrm{FFN}(\mathrm{LN}(\tilde{\boldsymbol{Z}})) + \tilde{\boldsymbol{Z}}.$$



3. Specformer –

- Eigenvalue Decoding

1. Construct multiple bases then combine.
2. *Learnable spectral filters*

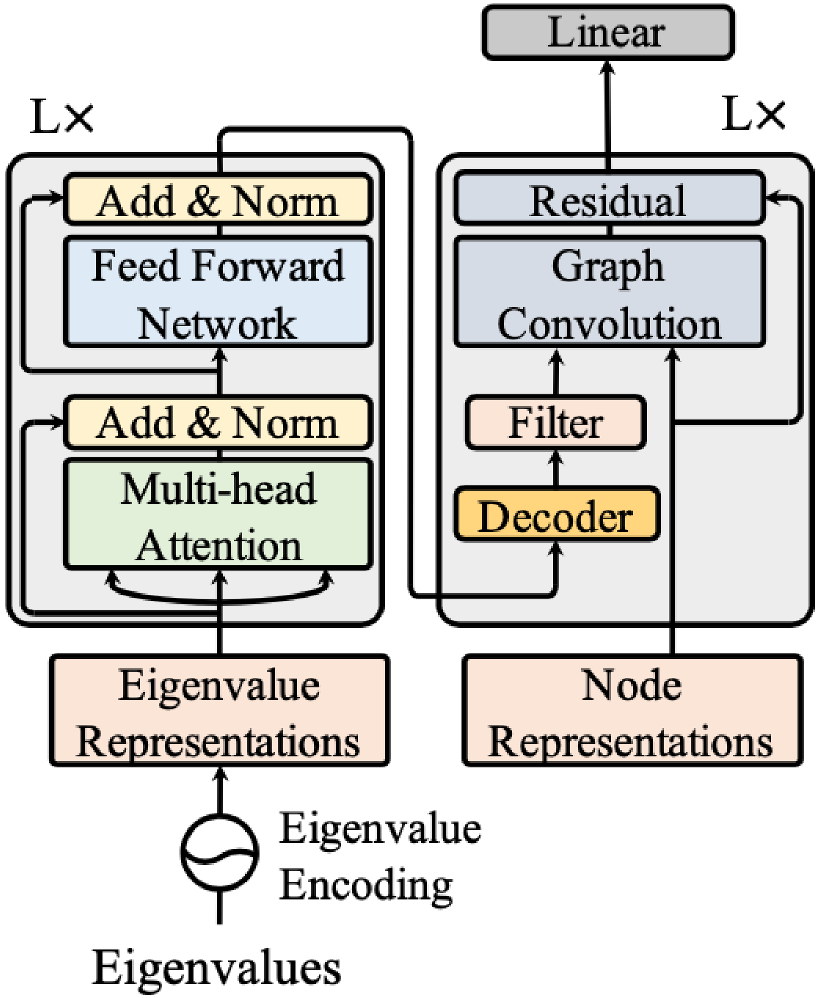
$$\boldsymbol{Z}_m = \text{Attention}(\boldsymbol{Q}\boldsymbol{W}_m^Q, \boldsymbol{K}\boldsymbol{W}_m^K, \boldsymbol{V}\boldsymbol{W}_m^V), \quad \boldsymbol{\lambda}_m = \phi(\boldsymbol{Z}_m \boldsymbol{W}_\lambda)$$

3. *Learnable bases*

$$\boldsymbol{S}_m = \boldsymbol{U}\text{diag}(\boldsymbol{\lambda}_m)\boldsymbol{U}^\top, \quad \hat{\boldsymbol{S}} = \text{FFN}([\boldsymbol{I}_n || \boldsymbol{S}_1 || \cdots || \boldsymbol{S}_M])$$

4. More customization during Convolution is possible:

   a. **Shared FFN and shared S_hat** (Specformer-small)
   b. **Layer-specific FFN and shared S_hat** (Specformer-medium)
   c. **Layer-specific FFN and layer-specific S_hat** (Specformer-large)



3. Specformer –

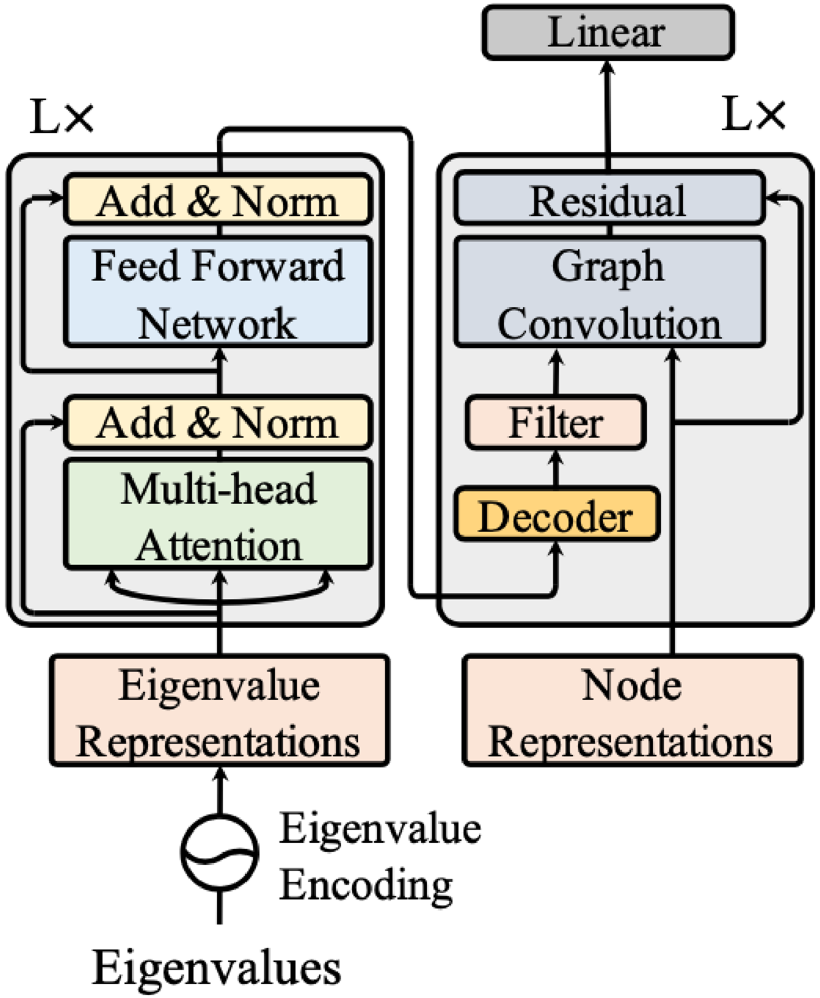
- Convolutions

$$\hat{\boldsymbol{X}}_{:,i}^{(l-1)} = \hat{\boldsymbol{S}}_{:,:,i}\boldsymbol{X}_{:,i}^{(l-1)}, \quad \boldsymbol{X}^{(l)} = \sigma\left(\hat{\boldsymbol{X}}^{(l-1)}\boldsymbol{W}_x^{(l-1)}\right) + \boldsymbol{X}^{(l-1)}$$

- Key Properties

  1. Vs **Polynomial GNNs**
  2. Vs **MPNNs**
  3. Vs **Graph Transformers**

- Can be scaled using sparse calculations

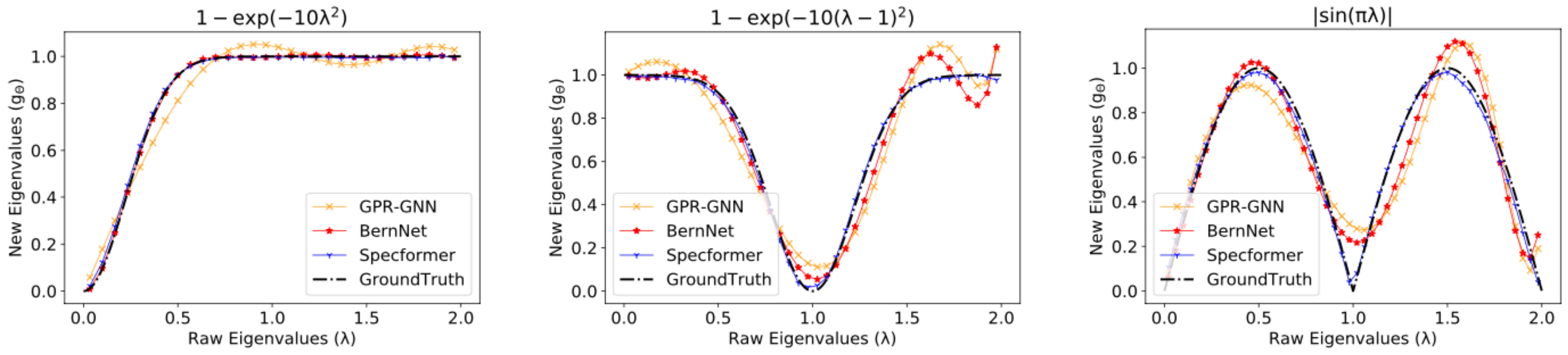


4. Experiments –

- Synthetic data

Table 1: Node regression results, mean of the sum of squared error & R^2 score, on synthetic data.

Model (~2k param.)	Low-pass $\exp(-10\lambda^2)$	High-pass $1-\exp(-10\lambda^2)$	Band-pass $\exp(-10(\lambda-1)^2)$	Band-rejection $1-\exp(-10(\lambda-1)^2)$	Comb $\lvert\sin(\pi\lambda)\rvert$
GCN	3.4799(.9872)	67.6635(.2364)	25.8755(.1148)	21.0747(.9438)	50.5120(.2977)
GAT	2.3574(.9905)	21.9618(.7529)	14.4326(.4823)	12.6384(.9652)	23.1813(.6957)
ChebyNet	0.8220(.9973)	0.7867(.9903)	2.2722(.9104)	2.5296(.9934)	4.0735(.9447)
GPR-GNN	0.4169(.9984)	0.0943(.9986)	3.5121(.8551)	3.7917(.9905)	4.6549(.9311)
BernNet	0.0314(.9999)	0.0113(.9999)	0.0411(.9984)	0.9313(.9973)	0.9982(.9868)
JacobiConv	0.0003(.9999)	0.0064(.9999)	0.0213(.9999)	0.0156(.9999)	0.2933(.9995)
Specformer	**0.0002(.9999)**	**0.0026(.9999)**	**0.0017(.9999)**	**0.0014(.9999)**	**0.0057(.9999)**



(a) High-pass (b) Band-rejection (c) Comb

Figure 2: Illustrations of filters learned by two polynomial GNNs and Specformer.

4. Experiments –

- Node classification data

Table 2: Results on real-world node classification tasks. Mean accuracy (%) ± 95% confidence interval. * means re-implemented baselines. "OOM" means out of GPU memory.

	Param. on Photo	Heterophilic				Homophilic			
		Chameleon	**Squirrel**	**Actor**	**Penn94**	**Cora**	**Citeseer**	**Photo**	**arXiv**
Spatial-based GNNs									
GCN	48K	59.61±2.21	46.78±0.87	33.23±1.16	82.47±0.27	87.14±1.01	79.86±0.67	88.26±0.73	71.74±0.29
GAT	49K	63.13±1.93	44.49±0.88	33.93±2.47	81.53±0.55	88.03±0.79	80.52±0.71	90.94±0.68	71.82±0.23
H_2GCN	60K	57.11±1.58	36.42±1.89	35.86±1.03	OOM	86.92±1.37	77.07±1.64	93.02±0.91	OOM
GCNII	49K	63.44±0.85	41.96±1.02	36.89±0.95	82.92±0.59	88.46±0.82	79.97±0.65	89.94±0.31	72.04±0.19
Spectral-based GNNs									
LanczosNet*	50K	64.81±1.56	48.64±1.77	38.16±0.91	81.55±0.26	87.77±1.45	80.05±1.65	93.21±0.85	71.46±0.39
ChebyNet	48K	59.28±1.25	40.55±0.42	37.61±0.89	81.09±0.33	86.67±0.82	79.11±0.75	93.77±0.32	71.12±0.22
GPR-GNN	48K	67.28±1.09	50.15±1.92	39.92±0.67	81.38±0.16	88.57±0.69	80.12±0.83	93.85±0.28	71.78±0.18
BernNet	48K	68.29±1.58	51.35±0.73	41.79±1.01	82.47±0.21	88.52±0.95	80.09±0.79	93.63±0.35	71.96±0.27
ChebNetII	48K	71.37±1.01	57.72±0.59	41.75±1.07	83.12±0.22	88.71±0.93	80.53±0.79	94.92±0.33	72.32±0.23
JacobiConv	48K	74.20±1.03	57.38±1.25	41.17±0.64	83.35±0.11	**88.98±0.46**	80.78±0.79	95.43±0.23	72.14±0.17
Graph Transformers									
Transformer*	37K	46.39±1.97	31.90±3.16	39.95±1.64	OOM	71.83±1.68	70.55±1.20	90.05±1.50	OOM
Graphormer*	139K	54.49±3.11	36.96±1.75	38.45±1.38	OOM	67.71±0.78	73.30±1.21	85.20±4.12	OOM
Specformer	32K	**74.72±1.29**	**64.64±0.81**	**41.93±1.04**	**84.32±0.32**	88.57±1.01	**81.49±0.94**	**95.48±0.32**	**72.37±0.18**

4. Experiments –

- Graph level tasks

Table 3: Results on graph-level datasets. ↓ means lower the better, and ↑ means higher the better.

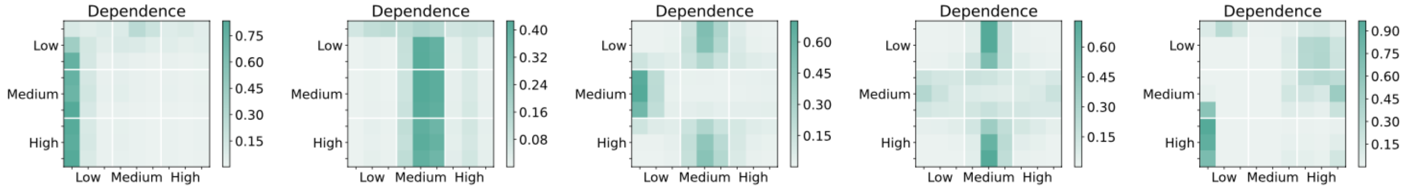
Model	ZINC(↓)	MolHIV(↑)	MolPCBA(↑)
GCN	0.367 ± 0.011	0.7599 ± 0.0119	0.2424 ± 0.0034
GIN	0.526 ± 0.051	0.7707 ± 0.0149	0.2703 ± 0.0023
GatedGCN	0.090 ± 0.001	-	0.267 ± 0.002
CIN	0.079 ± 0.006	**0.8094 ± 0.0057**	-
GIN-AK+	0.080 ± 0.001	0.7961 ± 0.0119	0.2930 ± 0.0044
GSN	0.101 ± 0.010	0.7799 ± 0.0100	-
DGN	0.168 ± 0.003	0.7970 ± 0.0097	0.2885 ± 0.0030
PNA	0.188 ± 0.004	0.7905 ± 0.0132	0.2838 ± 0.0035
Spec-GN	0.070 ± 0.002	-	0.2965 ± 0.0028
SAN	0.139 ± 0.006	0.7785 ± 0.0025	0.2765 ± 0.0042
Graphormer[2]	0.122 ± 0.006	0.7640 ± 0.0022	0.2643 ± 0.0017
GPS	0.070 ± 0.004	0.7880 ± 0.0101	0.2907 ± 0.0028
Specformer	**0.066 ± 0.003**	0.7889 ± 0.0124	**0.2972 ± 0.0023**

5. Ablations –

Table 4: Ablation studies on node-level and graph-level tasks.

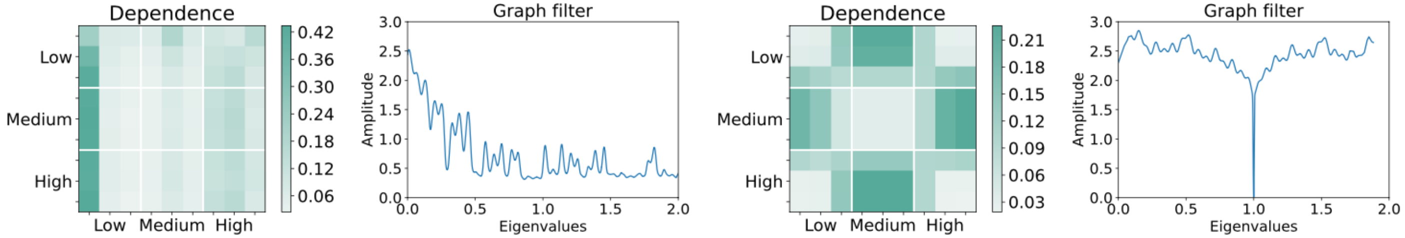
Encoder		**Decoder**			**Node-level**		**Graph-level**
$\rho(\lambda)$	**Attention**	**Small**	**Medium**	**Large**	**Squirrel (↑)**	**Citeseer (↑)**	**MolPCBA (↑)**
			✓		33.05	80.57	0.2696
✓			✓		63.78	81.17	0.2933
✓	✓		✓		64.64	81.49	0.2970
✓	✓	✓			64.51	81.47	0.2912
✓	✓			✓	65.10	80.00	0.2972

6. Visualizations(!!) –



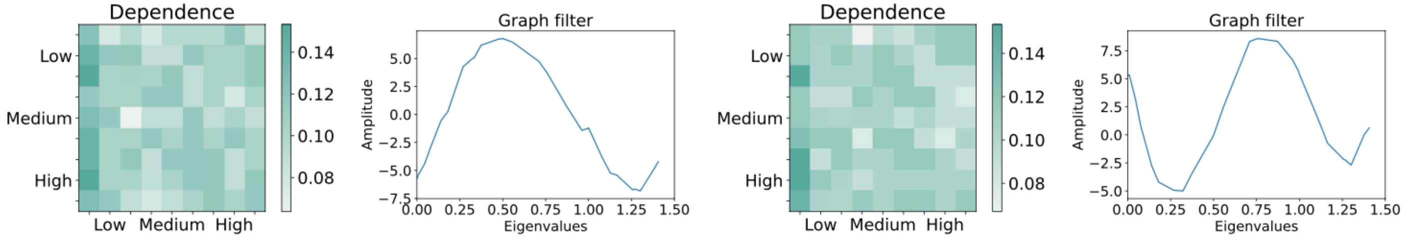
(a) Low-pass (b) High-pass (c) Band-pass (d) Band-rejection (e) Comb

Figure 3: The dependency of eigenvalues on synthetic graphs.



(a) Dependency of Citeseer (b) Filter of Citeseer (c) Dependency of Squirrel (d) Filter of Squirrel

Figure 4: The dependency and learned filters of heterophilic and homophilic datasets.



(a) dependency of $\hat{S}_1$ (b) Filter of $\hat{S}_1$ (c) dependency of $\hat{S}_2$ (d) Filter of $\hat{S}_2$

Figure 5: The dependency and basic filters of ZINC dataset.