GrokFormer: Graph Fourier Kolmogorov-Arnold Transformers

Guoguo Ai[1], Guansong Pang[2], Hezhe Qiao[2], Yuan Gao[1], Hui Yan[1]

[1]School of Computer Science and Engineering,
Nanjing University of Science and Technology, China

[2]School of Computing and Information Systems,
Singapore Management University, Singapore

Framework

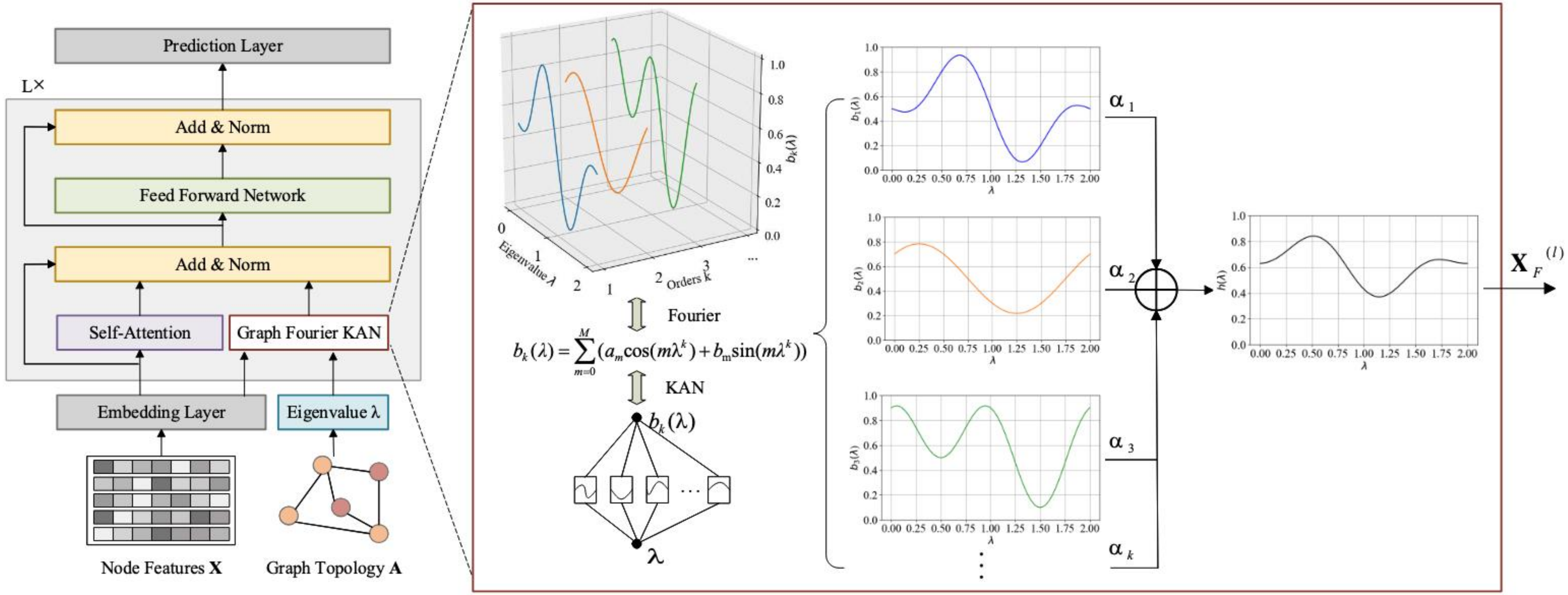

Figure 2: Overview of GrokFormer. In addition to the use of self-attention to capture global information in the spatial domain, a novel Graph Fourier KAN is proposed in GrokFormer to achieve global graph modeling in the spectral domain. This design enables a strong adaptability in both spectral order and graph spectrum, offering superior expressive power in capturing diverse graph frequency signals. GrokFormer synthesizes the spatial and spectral representations by a standard summation and normalization layer, followed by a Feed-Forward Network (FFN) layer for prediction.

The Proposed GrokFormer Filter

$$\phi_h(\lambda) = \sum_{k=1}^{K} \sum_{m=0}^{M} \left(\cos\left(m\lambda^k\right) \cdot a_{km} + \sin\left(m\lambda^k\right) \cdot b_{km}\right)$$

Order and Spectrum Adaptability

$$b_k(\lambda) = \sum_{m=0}^{M} \left(\cos\left(m\lambda^k\right) \cdot a_m + \sin\left(m\lambda^k\right) \cdot b_m\right) \qquad h(\lambda) = \sum_{k=1}^{K} \alpha_k b_k(\lambda)$$

Spectral Graph Convolution

$$\mathbf{X}_F^{(l)} = \mathbf{U}\,\mathrm{diag}(h(\lambda))\mathbf{U}^\top \mathbf{X}^{(l-1)}$$

Network Architecture of GrokFormer

$$\mathbf{X}'^{(l)} = EMHA\left(LN\left(\mathbf{X}^{(l-1)}\right)\right) + \mathbf{X}^{(l-1)} + \mathbf{X}_F^{(l)}$$

$$\mathbf{X}^{(l)} = FFN\left(LN\left(\mathbf{X}'^{(l)}\right)\right) + \mathbf{X}'^{(l)}$$

- Our graph filter is learnable in both spectral order and graph spectrum.

$$h(\lambda) = \sum_{k=1}^{K} \alpha_k \sum_{m=0}^{M} \left(\cos\left(m\lambda^k\right) \cdot a_{km} + \sin\left(m\lambda^k\right) \cdot b_{km}\right)$$

- Existing polynomial filters are a simplified variant of our graph filter.

$$h(\lambda) = \alpha_0 + \alpha_1\lambda + \alpha_2\lambda^2 + \cdots \alpha_K\lambda^K = \sum_{k=0}^{K} \alpha_k\lambda^k$$

- Specformer filter is a simplified variant of our graph filter.

$$h_s(\lambda) = a_0\lambda + \sum_{i=1}^{M} (\sin(m\lambda) \cdot a_i + \cos(m\lambda) \cdot b_i)$$

- Our filter $h(\lambda)$ can approximate any continuous function and constructs a permutation-equivariant spectral graph convolution.

Results

Table 2: Node classification results on five homophilic and five heterophilic datasets: mean accuracy (%) ± std. The best results are in bold, while the second-best ones are underlined. 'OOM' means out of memory

	Homophilic Datasets						Heterophilic Datasets				
	Cora	Citeseer	Pubmed	Photo	WikiCS	Physics	Penn94	Chameleon	Squirrel	Actor	Texas
Spatial-based GNNs											
GCN	87.14±1.01	79.86±0.67	86.74±0.27	88.26±0.73	82.32±0.69	97.74±0.35	82.47±0.27	59.61±2.21	46.78±0.87	33.23±1.16	77.38±3.28
GAT	88.03±0.79	80.52±0.71	87.04±0.24	90.94±0.68	83.22±0.78	97.82±0.28	81.53±0.55	63.13±1.93	44.49±0.88	33.93±2.47	80.82±2.13
H2GCN	87.96±0.37	80.90±1.21	89.18±0.28	95.45±0.67	83.45±0.26	97.19±0.13	81.31±0.60	61.20±4.28	39.53±0.88	36.31±2.58	91.89±3.93
HopGNN	88.68±1.06	80.38±0.68	89.15±0.35	94.49±0.33	84.73±0.59	97.86±0.16	OOM	65.25±3.49	57.83±2.11	39.33±2.79	89.15±4.04
Spectral-based GNNs											
ChebyNet	86.67±0.82	79.11±0.75	87.95±0.28	93.77±0.32	82.95±0.45	97.25±0.78	81.09±0.33	59.28±1.25	40.55±0.42	37.61±0.89	86.22±2.45
GPRGNN	88.57±0.69	80.12±0.83	88.46±0.33	93.85±0.28	82.58±0.89	97.25±0.13	81.38±0.16	67.28±1.09	50.15±1.92	39.92±0.67	92.95±1.31
BernNet	88.52±0.95	80.09±0.79	88.48±0.41	93.63±0.35	83.56±0.61	97.36±0.30	82.47±0.21	68.29±1.58	51.35±0.73	41.79±1.01	93.12±0.65
JacobiConv	88.98±0.46	80.78±0.79	89.62±0.41	95.43±0.23	84.13±0.49	97.56±0.28	83.35±0.11	74.20±1.03	57.38±1.25	41.17±0.64	93.44±2.13
HiGCN	89.23±0.23	81.12±0.28	89.95±0.13	95.33±0.37	83.14±0.78	97.65±0.35	OOM	68.47±0.45	51.86±0.42	41.81±0.52	92.15±0.73
Graph Transformers											
Transformer	71.83±1.68	70.55±1.20	86.66±0.50	89.58±1.05	77.36±1.25	OOM	OOM	45.21±2.01	33.17±1.32	39.95±0.64	88.75±6.30
GraphGPS	83.42±1.22	75.87±0.71	86.62±0.53	94.35±0.25	79.26±0.57	97.60±0.05	OOM	46.07±1.51	34.14±0.73	37.68±0.94	83.71±5.85
NodeFormer	87.32±0.92	79.56±1.10	89.24±0.23	95.27±0.22	81.03±0.94	96.45±0.28	69.66±0.83	56.34±1.11	43.42±1.62	34.62±1.82	84.63±3.47
SGFormer	87.87±2.67	79.62±1.63	89.07±0.14	94.34±0.23	82.71±0.56	97.96±0.81	76.65±0.49	61.44±1.37	45.82±2.17	41.69±0.63	92.46±1.48
NAGphormer	88.15±1.35	80.12±1.24	89.70±0.19	95.49±0.11	83.41±0.34	97.85±0.26	73.98±0.53	54.92±1.11	48.55±2.56	40.08±1.50	91.80±1.85
Specformer	88.57±1.01	81.49±0.94	90.61±0.23	95.48±0.32	85.15±0.63	97.75±0.53	**84.32±0.32**	74.72±1.29	64.64±0.81	41.93±1.04	88.23±0.38
PolyFormer	87.67±1.28	81.80±0.76	90.68±0.31	94.08±1.37	83.62±0.17	98.08±0.27	79.27±0.26	60.17±1.39	44.98±3.03	41.51±0.71	89.02±5.44
GrokFormer	**89.57±1.43**	**81.92±1.25**	**91.39±0.51**	**95.52±0.52**	**85.57±0.65**	**98.31±0.18**	83.59±0.26	**75.58±1.73**	**65.12±1.59**	**42.98±1.48**	**94.59±2.08**