Polynomial Neural Sheaf Diffusion: A Spectral Filtering Approach on Cellular Sheaves

Alessio Borgi, Fabrizio Silvestri, Pietro Liò

SAPIENZA UNIVERSITÀ DI ROMA

UNIVERSITY OF CAMBRIDGE

23/24 March 2026, University of Cambridge, United Kingdom

UNIVERSITY OF CAMBRIDGE

NEURASON'26 Workshop

Introduction & Background

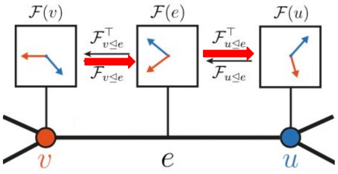
✗ PROBLEM

Most of GNNs present an accuracy degradation in heterophilic settings/when adding many layers:

- Rely on the **Homophilic Assumption**.
- Suffer from **Oversmoothing Effect**.

✅ SOLUTION → SHEAF NEURAL NETWORKS

- **False Assumption**: Graph Nodes live in the same global feature space.
- Assign a vector space (**stalk**) to each node and edge and learn **restriction maps** (and transport maps) among them.



How to move a vector x from v to u?

Moving $\mathbf{x} \in \mathcal{F}(v) \to \mathcal{F}(u)$ is done via maps composition!

$$\mathcal{F}^\top_{u\trianglelefteq e}\,\mathcal{F}_{v\trianglelefteq e}\,\mathbf{x} \in \mathcal{F}(u)$$

Sheaf Laplacian

Q: Can we measure the **total disagreement** at each node?

- Space of 0-Cochains: $C^0(\mathcal{G},\mathcal{F}) \coloneqq \bigoplus_{u\in\mathcal{V}} \mathcal{F}(u)$
- Space of 1-Cochains: $C^1(\mathcal{G},\mathcal{F}) \coloneqq \bigoplus_{e\in\mathcal{E}} \mathcal{F}(e)$
- Co-boundary Map: $\delta : C^0(\mathcal{G},\mathcal{F}) \to C^1(\mathcal{G},\mathcal{F})$
  - Measures the nodes disagreement

A: Build the **Sheaf Laplacian** $L_\mathcal{F} \coloneqq \delta^\top\delta$

$$L_\mathcal{F}(\mathbf{x})_u = \sum_{u,v\trianglelefteq e} \mathcal{F}^\top_{u\trianglelefteq e}\left(\mathcal{F}_{u\trianglelefteq e}\mathbf{x}_u - \mathcal{F}_{v\trianglelefteq e}\mathbf{x}_v\right)$$

Polynomial Neural Sheaf Diffusion

- Model-agnostic SheafNN framework to perform diffusion explicitly in the spectral domain.
- Applies Orthogonal Polynomials filters to the spectrally rescaled sheaf Laplacian.
- No need anymore for large stalk dimensions.
- Same asymptotic cost of stacking K NSD-layers but without Laplacian re-assembly.

Double Interpretation:

1. **Spectral**: Frequency–selective filter on the sheaf signal.
2. **Spatial**: The degree K determines the receptive field, mixing information across all paths of length up to K in a single pass.

$$p(\lambda) = \sum_{k=0}^{K} \theta_k B_k(\widetilde{L}), \qquad \widetilde{L} = \frac{2}{\lambda_{\max}} L - I$$

Research Question

Can we think Sheaf Diffusion under a Spectral Perspective, and if yes, can we use their properties for Sheaves?

Long-Range Influence

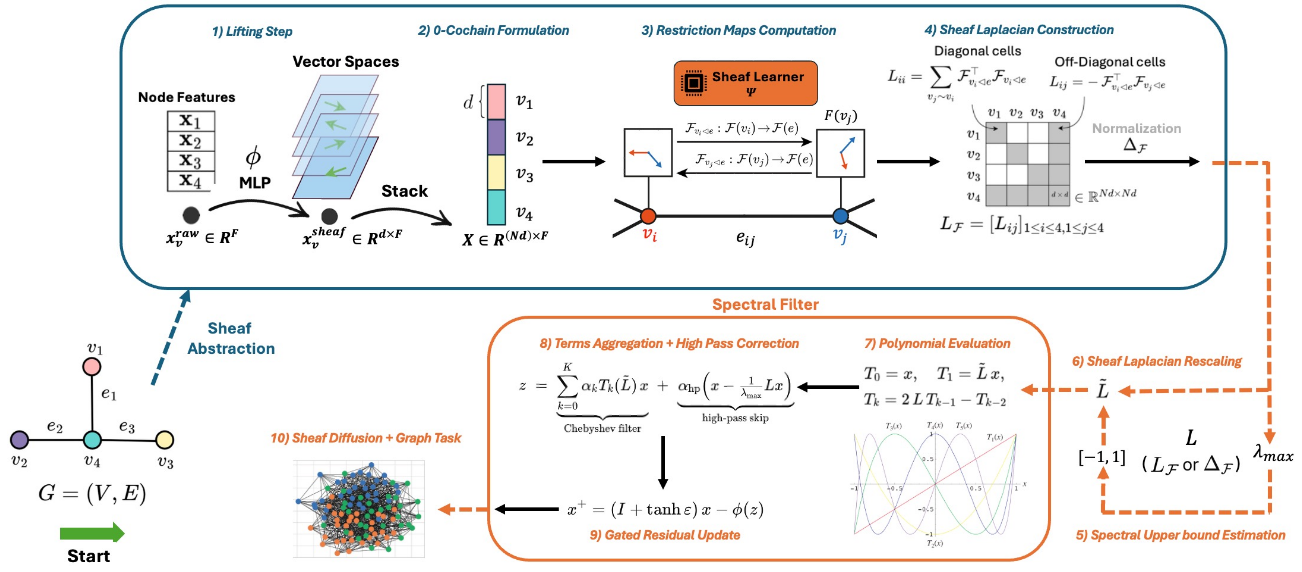
⚠ Long-range interactions require stacking many layers.
→ Increase computational cost and intensifying the oversmoothing.

Explicit K-hop receptive field/diffusion/long-range mixing within a single layer.

Controllable frequency-selective operator, enabling direct control over the diffusion behaviour (low-, band-, or high-pass).

→ Gradient-based Influence Decay to demonstrate persistent Long-Range influence.

$$I(d) \coloneqq \mathbb{E}_{v\in\mathcal{T}}\left[\frac{1}{|\mathcal{N}_d(v)|}\sum_{u\in\mathcal{N}_d(v)} G_{uv}\right]$$



Experiments & Results

Discrete Models

	Texas	Wisconsin	Film	Squirrel	Chameleon	Cornell	Citeseer	Pubmed	Cora
Homophily level	**0.11**	**0.21**	**0.22**	**0.22**	**0.23**	**0.30**	**0.74**	**0.80**	**0.81**
#Nodes	183	251	7,600	5,201	2,277	183	3,327	18,717	2,708
#Edges	295	466	26,752	198,493	31,421	280	4,676	44,327	5,278
#Classes	5	5	5	5	5	5	7	3	6
DiagPolySD	**90.00±4.68**	88.63±3.59	37.31±0.98	**56.61±2.06**	**71.45±2.03**	**86.49±5.54**	**77.74±1.26**	89.70±0.32	**88.79±1.13**
BundlePolySD	**89.74±5.32**	**89.41±4.04**	37.47±0.86	55.76±2.02	**71.18±1.46**	**86.76±4.90**	**77.57±1.55**	**89.75±0.34**	88.33±1.34
GeneralPolySD	**89.21±5.05**	88.82±4.89	37.34±1.13	**55.79±2.52**	69.62±1.85	**86.49±5.80**	77.21±1.58	89.73±0.41	**88.47±1.19**
PolySpectralGNN	64.59±6.40	58.62±6.04	25.50±0.85	51.72±1.76	62.65±3.03	54.59±6.26	60.4±0.89	78.7±0.56	76.2±0.67
RiSNN - NoT	87.89±4.28	88.04±2.39	*N/A*	51.24±1.71	66.58±1.81	82.97±6.17	75.07±1.56	87.91±0.55	85.86±1.31
RiSNN	86.84±3.72	87.84±2.60	*N/A*	53.30±3.30	65.15±2.40	85.95±6.14	76.23±1.81	88.00±0.42	85.27±1.11
JdSNN - NoW	87.30±4.53	88.43±2.83	*N/A*	51.28±1.80	66.45±3.46	84.59±6.95	75.93±1.41	88.09±0.49	84.39±1.47
JdSNN	87.37±5.10	**89.22±3.42**	*N/A*	49.89±1.71	66.40±2.33	85.41±4.55	73.27±1.86	88.19±0.55	85.43±1.73
Conn - NSD	86.16±2.24	88.73±4.47	**37.91±1.28**	45.19±1.57	65.21±2.04	85.95±7.72	75.61±1.93	89.28±0.38	83.74±2.19
SAN	84.05±5.33	86.47±3.87	37.09±1.18	50.96±1.40	67.46±1.90	84.32±5.64	72.57±1.50	87.12±0.30	85.90±1.85
ANSD	85.68±4.69	87.45±3.19	37.66±1.40	54.39±1.76	68.38±2.14	84.59±5.93	76.81±1.82	89.21±0.37	87.20±1.03
Diag - NSD	85.67±6.95	88.63±2.75	37.79±1.01	54.78±1.81	68.68±1.73	86.49±7.35	77.14±1.85	89.42±0.43	87.14±1.06
O(d) - NSD	85.95±5.51	**89.41±4.74**	**37.81±1.15**	**56.34±1.32**	68.04±1.58	84.86±4.71	76.70±1.57	89.49±0.40	86.90±1.13
Gen - NSD	82.97±5.13	89.21±3.84	**37.80±1.22**	53.17±1.31	67.93±1.58	85.68±6.51	76.32±1.65	89.33±0.35	87.30±1.15
GGCN	84.86±4.55	86.86±3.29	37.54±1.56	55.17±1.58	**71.14±1.84**	85.68±6.63	77.14±1.45	89.15±0.37	87.95±1.05
H2GCN	84.86±7.23	87.65±4.98	35.70±1.00	36.48±1.86	60.11±2.15	82.70±5.28	77.11±1.57	89.49±0.38	87.87±1.20
GPRGNN	78.38±4.36	82.94±4.21	34.63±1.22	31.61±1.24	46.58±1.71	80.27±8.11	77.13±1.67	87.54±0.38	87.95±1.18
FAGCN	82.43±6.89	82.94±7.95	34.87±1.25	42.59±0.79	55.22±3.19	79.19±9.79	*N/A*	*N/A*	*N/A*
MixHop	77.84±7.73	75.88±4.90	32.22±2.34	48.30±1.48	60.50±2.53	73.51±6.34	76.26±1.33	85.31±0.61	87.61±0.85
GCNII	77.57±3.83	80.39±3.40	37.44±1.30	38.47±1.58	63.86±3.04	77.86±3.79	77.33±1.48	**90.15±0.43**	**88.37±1.25**
Geom - GCN	66.76±2.72	64.51±3.66	31.59±1.15	38.15±0.92	60.00±2.81	60.54±3.67	**78.02±1.15**	**89.95±0.47**	85.35±1.57
PairNorm	60.27±4.34	48.43±6.14	27.40±1.24	50.44±2.04	62.74±2.82	58.92±3.15	73.59±1.47	87.53±0.44	85.79±1.01
GraphSAGE	82.43±6.14	81.18±5.56	34.23±0.99	41.61±0.74	58.73±1.68	75.95±5.01	76.04±1.30	88.45±0.50	86.90±1.04
GCN	55.14±5.16	51.76±3.06	27.32±1.10	53.43±2.01	64.82±2.24	60.54±5.30	76.50±1.36	88.42±0.50	86.98±1.27
GAT	52.16±6.63	49.41±4.09	27.44±0.89	40.72±1.55	60.26±2.50	61.89±5.05	76.55±1.23	87.30±1.10	86.33±0.48
MLP	80.81±4.75	85.29±3.31	36.53±0.70	28.77±1.56	46.21±2.99	81.89±6.40	74.02±1.90	75.69±2.00	87.16±0.37

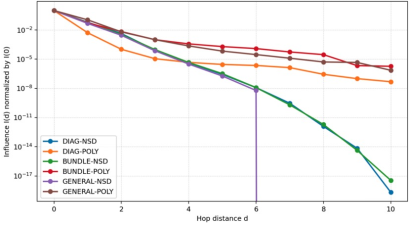
- ▶ New s.o.t.a. results on heterophilic and homophilic datasets.
- ▶ Diag models invert NSD* trend → performances speed-up.
- ▶ Stalk dimension has not big impact as it was for NSD.
- ▶ No Oversmoothing →High results if L increases.
- ▶ Same is valid for the Continuous Neural Sheaf Diffusion.

Continuous Models

	Texas	Wisconsin	Film	Squirrel	Chameleon	Cornell	Citeseer	PubMed	Cora
Homophily	**0.11**	**0.21**	**0.22**	**0.22**	**0.23**	**0.30**	**0.74**	**0.80**	**0.81**
#Nodes	183	251	7,600	5,201	2,277	183	3,327	18,717	2,708
#Edges	295	466	26,752	198,493	31,421	280	4,676	44,327	5,278
#Classes	5	5	5	5	5	5	7	3	6
Cont DiagChebySD	**86.58±5.45**	**88.04±3.45**	36.80±1.01	**53.76±1.78**	**65.74±3.50**	**84.86±5.30**	**76.71±1.63**	**89.58±0.46**	**87.83±1.19**
Cont BundleChebySD	**86.84±7.16**	86.47±6.17	**37.42±1.12**	52.34±2.43	**66.49±1.55**	84.59±5.27	75.66±1.60	**89.70±0.31**	87.73±1.26
Cont GeneralChebySD	83.42±5.53	**86.86±3.40**	36.76±0.96	**52.87±1.89**	64.10±2.35	**84.86±4.22**	76.01±2.03	89.32±0.38	**87.79±1.34**
Cont Diag-NSD	82.97±4.37	**86.47±2.55**	**36.85±1.21**	38.17±9.29	62.06±3.84	80.06±6.07	76.56±1.19	89.47±0.42	86.88±1.21
Cont O(d)-NSD	82.43±5.95	84.50±4.34	36.39±1.37	40.40±2.01	63.18±1.69	72.16±10.40	75.19±1.67	89.12±0.30	86.70±1.24
Cont Gen-NSD	**83.78±6.62**	85.29±3.31	**37.28±0.74**	**52.57±2.82**	**66.40±2.28**	84.60±4.68	**77.54±1.72**	**89.67±0.40**	87.45±0.99
BLEND	83.24±4.65	84.12±3.58	35.63±0.89	43.06±1.39	60.11±2.09	**85.95±6.82**	76.63±1.60	89.24±0.42	**88.09±1.22**
GRAND	75.68±7.25	79.41±3.64	35.62±1.01	40.05±1.50	54.67±2.54	82.16±7.09	76.46±1.77	89.02±0.52	87.36±0.96
CGNN	71.35±4.05	74.31±7.26	35.95±0.86	29.24±1.09	46.89±1.69	66.22±7.62	**76.91±1.81**	87.70±0.49	87.10±1.35

Long-Range Impact - Influence Decay

SQUIRREL | Influence decay vs hop distance (normalized) | log-scale



PolyNSD VS NSD – #Layers Ablation

Layers	2	4	8	16	32	Best	PolySD Improvement
PubMed (h=0.80, $\#N = 18{,}707$, $\#E = 44{,}327$, $\#C = 3$) **DiagChebySD**: 88.18±0.40 (K=8), #params=48,665 **BundleChebySD**: 87.91±0.32 (K=4), #params=49,621 **GeneralChebySD**: 87.87±0.49 (K=8), #params=52,505							
Diag-NSD	87.82±0.55	**87.92±0.51**	87.92±0.52	65.92±20.39	39.49±1.60	4	**+0.26%**
#params	48,655	50,507	54,211	61,619	76,435	50,507	-3,65% (-1,842)
Bundle-NSD	87.70±0.56	87.85±0.42	**87.94±0.36**	87.63±0.47	37.03±4.72	8	**-0.03%**
#params	49,615	52,427	58,051	69,299	91,795	58,051	-14,52%(-8,430)
General-NSD	87.48±0.64	87.62±0.36	**87.72±0.68**	39.94±1.03	39.33±2.25	8	**+0.15%**
#params	52,495	58,187	69,571	92,339	137,875	69,571	-24,53%(-17,066)
Chameleon (h=0.23, $\#N = 2{,}277$, $\#E = 31{,}421$, $\#C = 5$) **DiagChebySD**: 70.57±1.32 (K=16), #params=194,835 **BundleChebySD**: 66.64±2.35 (K=8), #params=195,787 **GeneralChebySD**: 67.39±2.50 (K=16), #params=198,675							
Diag-NSD	**64.43±2.06**	61.27±5.14	57.34±6.10	22.92±1.42	22.89±2.59	2	**+6.15%**
#params	194,817	196,669	200,373	207,781	222,597	194,817	+0.009%(+18)
Bundle-NSD	47.10±10.14	**54.08±6.30**	50.57±3.53	24.91±2.87	23.05±2.46	4	**+12.56%**
#params	195,777	198,589	204,213	215,461	237,957	198,589	-1,41%(-2,802)
General-NSD	**59.60±4.53**	58.18±3.56	26.05±3.41	23.79±3.41	20.07±3.74	2	**+7.79%**
#params	198,657	204,349	215,733	238,501	284,037	198,657	+0%(+0)
Squirrel (h=0.22, $\#N = 5{,}201$, $\#E = 198{,}493$, $\#C = 5$) **DiagChebySD**: 47.72±1.20 (K=16), #params=174,675 **BundleChebySD**: 44.48±3.22 (K=16), #params=176,913 **GeneralChebySD**: 42.02±2.50 (K=6), #params=179,785							
Diag-NSD	41.98±1.17	**42.60±1.83**	42.05±1.54	20.80±1.33	OOM	4	**+5.12%**
#params	175,937	177,789	181,493	188,901	N/A	177,789	-1,75%(-3,114)
Bundle-NSD	**42.46±1.45**	42.30±1.83	37.99±2.61	22.86±3.03	OOM	2	**+2.02%**
#params	176,897	179,709	185,333	196,581	N/A	176,897	+0.009%(+16)
General-NSD	39.11±1.96	**39.96±1.76**	33.69±1.32	OOM	OOM	4	**+2.06%**
#params	179,777	185,469	196,853	N/A	N/A	185,469	-3.06%(-5,684)

- ▶ Pick best K for the three models and run #layers and #channels ablation.
- ▶ PolyNSD has low K for homophilic and high K for heterophilic datasets.
- ▶ Even if more layers/hidden channels, PolyNSD is better → Parameter Savings.

---

*C. Bodnar, F. Di Giovanni, B. Chamberlain, P. Lio, and M. Bronstein. Neural Sheaf Diffusion: A topological perspective on heterophily and oversmoothing in GNNs. *NeurIPS*, 2022.