A Unified Uncertainty Representation for Graph Neural Networks via Doubly-Spectral Stochastic Expansion

Fred Xu[1,*], Thomas Markovich[2], Florence Regol[3], Yizhou Sun[1]

[1]UCLA [2]Block, Inc. [3]Mila, Quebec AI Institute [*]Work done during an internship at Block, Inc.

UCLA Samueli Computer Science

block

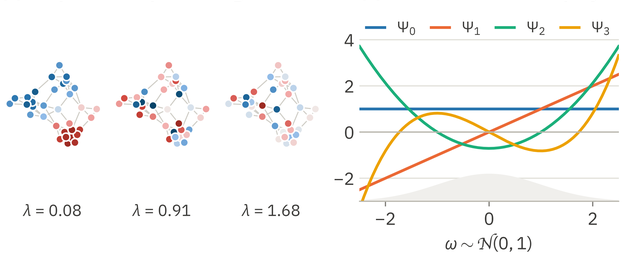
Motivation

- Deployed GNNs (drug discovery, fraud detection) need **calibration** (confidence matches accuracy), **OOD detection** (low confidence outside the training support) and **shift robustness** (accuracy degrades slowly); on graphs, uncertainty depends on the neighborhood as well as the features
- **Existing methods each target one**: sampling costs M forward passes; GPN lacks shift-robust classification; energy scores are not calibrated probabilities

Two Spectral Axes

- **Graph frequency**: Laplacian eigenvectors u_j (eigenvalue λ_j) form an orthonormal basis of graph signals $\ell^2(V)$; Chebyshev filters $g(L) = \sum_k c_k T_k(\tilde{L})$ of the rescaled Laplacian $\tilde{L}$ can be low- or high-pass
- **Chaos order**: normalized Hermite polynomials $\Psi_n = \mathrm{He}_n/\sqrt{n!}$ form an orthonormal basis of $L^2(\mathbb{R}, \gamma)$, $\gamma = \mathcal{N}(0,1)$, eigenfunctions of the Ornstein-Uhlenbeck operator ($-\mathcal{L}\Psi_n = n\Psi_n$)

(a) Graph axis: Laplacian eigenvectors **(b) Chaos axis: Hermite polynomials**



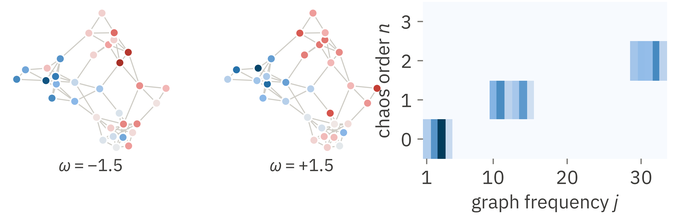
- **Polynomial chaos**: $Y(\omega) = \sum_n Y_n \Psi_n(\omega)$ has mean Y_0 and variance $\sum_{n\geq 1} Y_n^2$ (the **chaos energy**)

Doubly-Spectral Expansion

- A random graph signal driven by one shared Gaussian ω lies in $\ell^2(V) \otimes L^2(\mathbb{R}, \gamma)$, with complete orthonormal basis $\{u_j \otimes \Psi_n\}$; $\hat{Y}_{n,j}$ is the coefficient at graph frequency λ_j and chaos order n:

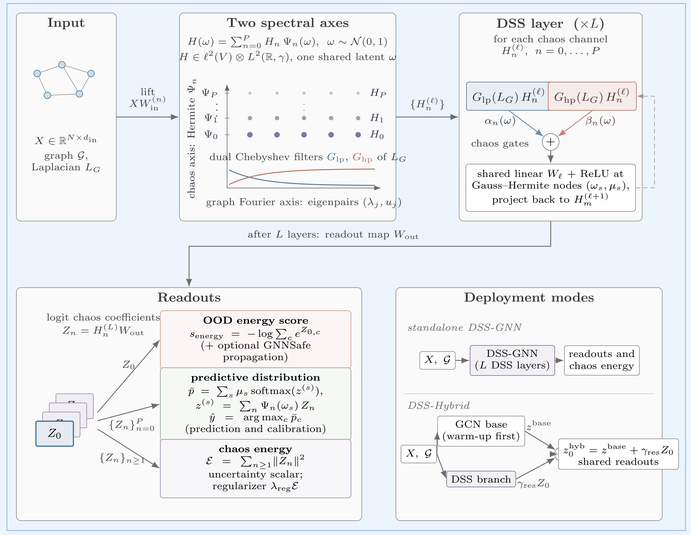
$$Y(\omega) = \sum_{n\geq 0} \sum_{j=1}^{N} u_j \hat{Y}_{n,j}^{\top} \Psi_n(\omega)$$

(a) Two outcomes of $Y(\omega)$ **(b) $|\hat{Y}_{n,j}|$ on the two axes**



DSS-GNN: One Uncertainty Representation, Task-Matched Readouts

TL;DR: DSS-GNN stores P+1 chaos coefficients per node, filters them along graph frequency and chaos order, and integrates out the Gaussian latent by quadrature: one deterministic pass, no sampling, with task-matched readouts



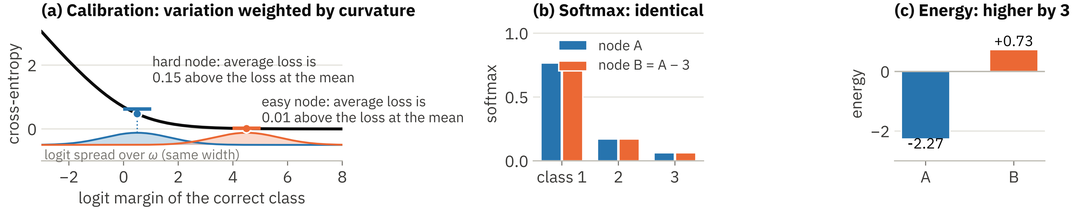
State: $H^{(\ell)}(\omega) = \sum_{n=0}^{P} H_n^{(\ell)} \Psi_n(\omega)$
Graph axis: each H_n goes through low- and high-pass Chebyshev filters $G_{\mathrm{lp}}(L)$, $G_{\mathrm{hp}}(L)$
Chaos axis: per-order gates $\alpha_n(\omega), \beta_n(\omega)$ combine the two branches
Nonlinearity: evaluate at S Gauss-Hermite nodes, apply W_ℓ and ReLU σ, project back

$$H_m^{(\ell+1)} = \sum_{s=1}^{S} \mu_s \, \sigma(\tilde{H}^{(\ell+1)}(\omega_s)) \Psi_m(\omega_s)$$

Quadrature: $S = 4$ is exact for the filter and gate terms at $P \leq 3$ (the recommended $P = 1$ to 2); the ReLU projection is approximate
Training: cross-entropy of the mean logit $Z_{i,0}$ plus $\lambda_{\mathrm{reg}} \mathcal{E}_i$, with chaos energy $\mathcal{E}_i = \sum_{n\geq 1} \|Z_{i,n}\|^2$
Two modes: standalone, or DSS-Hybrid: a DSS branch beside a warmed-up GCN, $z_i^{\mathrm{hyb}} = z_i^{\mathrm{base}} + \gamma_{\mathrm{res}} Z_{i,0}$
Guidance: standalone for calibration and heterophily; hybrid with a strong base encoder, for cross-graph and sparse-label OOD

Readouts from One Expansion

(a) Calibration: variation weighted by curvature **(b) Softmax: identical** **(c) Energy: higher by 3**



- **Predictive**: $\bar{p}_i = \sum_s \mu_s \, \mathrm{softmax}(z_i^{(s)})$ with $z_i^{(s)} = \sum_n Z_{i,n} \Psi_n(\omega_s)$; prediction $\arg\max \bar{p}_i$ and calibration use one distribution
- **Energy score**: $s(i) = -\log \sum_c e^{Z_{i,0,c}}$ from the mean logit, with GNNSafe score propagation
- **Prop. 2** ($S \geq P$+1): (a) $\tilde{\ell}_i = \ell(Z_{i,0}, y_i) + \frac{1}{2}\mathrm{tr}(\mathcal{H}_i \Sigma_{z_i}) + O(\cdot)$: chaos covariance weighted by loss curvature, $\leq \frac{1}{4}\mathcal{E}_i$ for cross-entropy; (b) a common logit shift leaves the softmax unchanged but moves the energy by exactly that shift
- **Readouts agree**: $\arg\max \bar{p}_i$ and $\arg\max Z_{i,0}$ differ on at most 0.21% of test nodes (none on 10 of 14 datasets)

Standalone vs. Hybrid, Cross-Evaluated

Task (settings)	Standalone DSS-GNN	DSS-Hybrid
Calibration (14)	Best Brier on 14; beats the hybrid on 13	No loss to its GCN base beyond noise
OOD (11)	Effective on 10; not on Twitch	Beats the standalone on 10
GOOD (7)	Above all baselines on 5, ERM level on 2	Best on all 7

Theory

- **Generalization** (Prop. 1): with $P = 0$ and scalar gates, a DSS layer reduces to a polynomial spectral GNN layer (ChebNet, GCN)
- **Doubly-spectral operator** (Thm. 1): without the nonlinearity and with exact quadrature, a DSS layer acts on the joint domain frequency by frequency: filters set the graph response, gates the chaos-order mixing
- **Capacity** (Thm. 2): with $\mathrm{rank}(X) = N$, a one-layer subfamily matches the chaos coefficients of any Gaussian-latent random graph signal through order P, with truncation error $\leq C\beta^P$ ($\beta < 1$) for targets with an entire Gaussian extension of controlled growth; for $\mathrm{rank}(X) < N$, it matches every truncated target whose coefficients lie in $\mathrm{range}(X)$, under one positivity condition
- These are capacity results; the task-level claims are empirical

Calibration (Brier ↓, 10 splits)

Dataset	Base	Ours	Dataset	Base	Ours
Cora	0.226	**0.207**	Squirrel	0.395	**0.394**
Citeseer	0.433	**0.308**	Coauthor-CS	0.082	**0.062**
PubMed	0.179	**0.156**	Roman-Emp.	0.414	**0.300**
Texas	0.255	**0.240**	Amazon-Rat.	0.637	**0.634**
Cornell	0.290	**0.231**	Minesweeper	0.248	**0.169**
Wisconsin	0.292	**0.104**	Tolokers	0.289	**0.268**
Chameleon	0.365	**0.339**	Questions	0.054	**0.053**

- Base: the better of TFE-GNN and G-ΔUQ on each dataset. **Lowest Brier on 14 of 14**, best accuracy on 13, no post-hoc calibration
- **Chaos order**: $P > 0$ lowers Brier on 12 of 14; $P = 1$ or 2 is within 0.02 Brier of the best order on every dataset

OOD Detection (AUROC, %)

DSS-Hybrid	Structure	Feature	Label
Cora	**94.32** (+3.70)	**97.60** (+2.04)	**94.11** (+0.36)
Photo	99.69 (−0.13)	**99.66** (+0.02)	**97.52** (+0.01)
CS	99.96 (−0.03)	99.95 (−0.02)	**98.07** (+0.16)
Cross-graph	Twitch **95.75** (+0.39)		Arxiv 73.36 (−1.41)

- In parentheses: difference from the best of 6 baselines. **Best on 6 of 9** node-level settings, within 0.13 on the rest; above MC-dropout and Deep Ensembles on all 11, in a single forward pass
- Same readout, objective and propagation as GNNSafe++, so the Cora gains come from the DSS branch; AUROC is flat in P (separate runs, Cora-structure, $P = 0$ to 3: 91.0/91.2/90.9/90.8)

Concept Shift (GOOD, accuracy %)

	CBAS	WebKB	Twitch	Cora-w	Cora-d	Arx-t	Arx-d
Best of 13	87.29	30.83	57.20	64.73	61.73	66.08	64.34
DSS-Hybrid	**88.57**	**40.77**	**61.21**	**64.81**	**62.73**	**66.76**	**66.43**

- Cora-w/d: word/degree splits; Arx-t/d: time/degree. Plain ERM training against invariance and robust-training baselines (TAR best on 6): **best on all 7**, by 0.08 to 9.94 points