GRAPH NONLINEAR SCHRÖDINGER NETWORKS FOR ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Supervised graph anomaly detection has to serve two frequency regimes at once because camouflaged fraudsters mimic their neighbors while other anomalies clash with their neighborhoods. Any single spectral response therefore fails in some regime. We introduce G-SNLS, a deterministic graph network whose depth is a bank of Schrödinger-type propagators. Each head learns its own damping, dispersion and time scale so that attenuation and phase velocity separate while several regimes coexist. Intensity-dependent coupling makes the flow nonlinear while paired decoders absorb the class imbalance. We prove an exact energy-decay law per step together with closure of the response family under composition and a full-depth state bound. G-SNLS takes the leading average rank on both ranking metrics across eight public benchmarks. Through ablations, we attribute the gains to the intensity coupling and the channel map. Compared with an unconstrained filter of matched degree, the constraint preserves accuracy while it yields a response that cannot diverge with depth and a loss surface whose most curved direction is flatter per unit change in the response.

1 INTRODUCTION

Payment networks carry money laundering, marketplaces carry fake reviews and social platforms carry accounts that are later banned. Supervised graph anomaly detection casts each of these as node classification on an interaction graph, and two properties of the task undermine standard graph learning methods. Positives are scarce and often amount to a few percent of the nodes, so accuracy-driven objectives and unweighted losses collapse onto the majority class. Anomalies also do not share one local structure. Fraudsters camouflage themselves by imitating benign neighbors so that their features end up smoother than the surrounding population (Dou et al., 2020). Banned users and planted outliers do the opposite and clash with their neighborhoods (Tang et al., 2022; Gao et al., 2023). A model restricted to one frequency response therefore fails on some datasets and occasionally on individual nodes of a single dataset.

Generic backbones fix one response in advance, whether it is the low-pass operator of graph convolution (Kipf & Welling, 2017; Wu et al., 2019; NT & Maehara, 2019), the ego term that sampling-based aggregation keeps beside the neighborhood summary (Hamilton et al., 2017), the signed polynomial of a learnable hop filter (Chien et al., 2021) or the diffusion equation that carries the smoothing of stacked convolution into continuous time (Chamberlain et al., 2021; Li et al., 2018; Oono & Suzuki, 2020). Anomaly-specific spectral architectures add fixed beta-wavelet bands (Tang et al., 2022), per-node attention over a node-invariant Bernstein pair (Chai et al., 2022), spectrally guided edge pruning (Gao et al., 2023), fused frequency branches (Liu et al., 2026b) or an explicit model of energy redistribution (Liu et al., 2026a).

Each of these fixes its response before training. A band structure is computed in advance or an edge-pruning decision is taken once, and attention runs over channels that freeze when training ends. We argue instead that a detector should carry several dynamical regimes at the same time and should learn the rate of each one rather than choose among a fixed set. Each head carries four physical parameters whereas a free polynomial filter of the same degree needs K complex coefficients per head and layer. Such a filter reaches the same accuracy when it starts from our responses (Section 4.4), while under our parameterization the most curved direction of the loss is flatter per unit change in the response and no width or depth can make the response diverge.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

G-SNLS, the graph spectral nonlinear Schrödinger network proposed here, splits its channels into spectral heads so that each head can learn its own damping, dispersion and time scale. Every layer evolves the whole state with one grouped spectral propagator inspired by damped Schrödinger dynamics (Sulem & Sulem, 2007). Dispersion distinguishes this flow from diffusion. A head can rotate the phase of a frequency band without attenuating it, so transport and smoothing become two learned quantities instead of two extremes of a single parameter. A deterministic nonlinear modulation then sits between propagation steps and writes local intensity and node content into phase and amplitude. Two decoders read the evolved state afterwards. One places the decision boundary under focal modulation while the other orders the score list under a smooth average-precision surrogate. We establish an exact spectral energy law together with a depth profile that keeps the composed response inside the family, a full-depth bound and permutation equivariance.

In Section 4.3 we attribute the largest effects of the study to the dissipative intensity coupling and to the complex channel map that couples the heads. These two terms improve accuracy whereas the constrained response flattens the most curved direction of the loss.

Our contributions are:

- We introduce a bank of nonlinear spectral dynamics whose damping, dispersion and time scale are learned per head and coupled between steps by an intensity-dependent modulation. We prove an exact spectral energy law, a depth profile that holds the composed response contractive at every frequency and exact on the lowest one, a full-depth state bound and node-relabeling equivariance.
- We pair deterministic multi-scale graph context with a two-headed objective that trains one decoder under a focal loss and the other under a differentiable average-precision surrogate at fixed weights. The state crossing a layer stays real, and one complex channel map is shared across depth. A component study measures the contribution of each part on six datasets.
- We show that the constrained response, with four physical parameters per head in place of K complex coefficients per head and layer, matches the accuracy of an unconstrained filter of matched degree. The most curved direction of its loss stays flatter per unit change in the response on every dataset we measure, and no width or depth can make the response diverge.

2 Related Work

Graph Anomaly Detection. Tang et al. (2022) show that anomalies shift graph signal energy toward higher frequencies and develop BWGNN from fixed beta-wavelet band-pass kernels. Gao et al. (2023) prune inter-class edges found in the high-frequency component of the label signal while Chai et al. (2022) learn to weight low-pass and high-pass channels per node. Recent detectors make the frequency response itself adaptive. SAGAD splits propagation into complementary frequency branches and fuses them to handle homophily disparity (Liu et al., 2026b) while EGNN models the spectral energy redistribution that anomalies induce (Liu et al., 2026a). A parallel line addresses the same task from the label side rather than the spectral one, either by resampling neighborhoods so that a scarce positive class survives aggregation (Liu et al., 2021) or by learning which homophily a node should be read under (Ai et al., 2025). We compare against the five spectral detectors BWGNN, GHRN, AMNet, SAGAD and EGNN. AMNet sets its per-node weights from the input features and applies them to a shared Bernstein filter pair. The channels of G-SNLS are instead learned nonlinear dissipative flows, and no band structure or pruning step is fixed before training.

Spectral and Dynamical Graph Networks. The response of GCN is low-pass (Kipf & Welling, 2017) while the separate self and neighbor transforms of GraphSAGE keep a term that propagation cannot smooth away (Hamilton et al., 2017). Chebyshev networks approximate functions of the graph Laplacian without an eigendecomposition (Defferrard et al., 2016) and GPR-GNN learns signed polynomial responses (Chien et al., 2021). BernNet constrains the response with Bernstein bases (He et al., 2021) and supplies the basis from which AMNet constructs its pair, while a low-pass and high-pass pair mixed per node adapts the response to heterophily (Bo et al., 2021; Zhu et al., 2020). A second family treats depth as time and takes its inductive bias from a partial differential equation. GRAND casts message passing as continuous diffusion (Chamberlain et al., 2021) and PDE-GCN derives diffusive and hyperbolic architectures from the same principle (Eliasof et al., 2021), while

GraphCON couples oscillators to resist oversmoothing (Rusch et al., 2022). Reaction terms restore the high-frequency content that pure diffusion destroys, either through a reaction-diffusion right-hand side (Choi et al., 2023) or through attractive and repulsive Allen–Cahn forces (Wang et al., 2023). Complex-valued propagation appears in magnetic-Laplacian networks for directed graphs where phase encodes structural asymmetry (Zhang et al., 2021; Fiorini et al., 2023). The governing equation of G-SNLS is dispersive instead of parabolic or hyperbolic, so a head can transport a frequency band without attenuating it. G-SNLS also runs a bank of such flows that a nonlinear modulation couples to node content between steps.

3 METHOD

Figure 1 shows the pipeline. Deterministic graph context is encoded into a complex state. A bank of grouped spectral propagators then evolves that state with a nonlinear modulation between the steps, and a decoder reads the result.

3.1 GRAPH CONTEXT AND ENCODING

Let $G = (\mathcal{V}, \mathcal{E})$ be an undirected graph with N nodes, adjacency matrix A, degree matrix D and feature matrix $X \in \mathbb{R}^{N\times F}$. Its symmetric normalized adjacency and Laplacian are

$$\widehat{A} = D^{-1/2}AD^{-1/2} + I_0, \qquad L = I - \widehat{A} = U\Lambda U^\top, \qquad \operatorname{spec}(L) \subseteq [0, 2], \tag{1}$$

with U orthogonal and Λ the diagonal matrix of eigenvalues. The diagonal I_0 marks the isolated nodes where $D^{-1/2}$ vanishes, so that each lies in the Laplacian nullspace as in Chung (1997). Before any learned computation runs, we expose graph context that a spectral evolution cannot cheaply recover on its own. The richest variant is the multi-scale context

$$X_1 = \widehat{A}X, \quad X_2 = \widehat{A}^2X, \quad R = |X - X_1|, \quad Q = \left[\widehat{A}(X \odot X) - X_1 \odot X_1\right]_+^{1/2},$$
$$C(X) = [X, X_1, X_2, R, Q, X \odot X_1], \tag{2}$$

where $\odot$ is the elementwise product and $[\cdot]_+$ clamps at a small positive floor. Here X_1 and X_2 summarize one-hop and two-hop neighborhoods while R measures feature disagreement and $X \odot X_1$ captures feature agreement. The term Q is a local-variation proxy whose clamp keeps it well-defined when symmetric normalization makes a row sum exceed one. We keep three smaller variants: the four-term homophily context $[X, X_1, R, X \odot X_1]$, the neighbor context $[X, X_1]$ and the raw input X. We pick the variant per dataset on validation data. No variant opens a direct input-to-output bypass because every coordinate of $C(X)$ passes through the full evolution described next. A linear encoder E forms the initial state

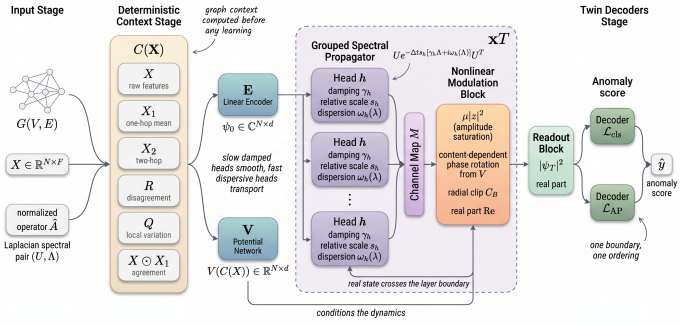

Figure 1: G-SNLS architecture. Deterministic multi-scale context is encoded as a complex state. The dashed frame is one layer that runs T times. Its grouped spectral propagator evolves every head once while the nonlinear modulation mixes channels and writes local intensity into phase and amplitude. The state crossing a layer boundary is real, and paired decoders read the final state under a focal loss and a ranking surrogate.

$$\psi_0 = E(C(X)) \in \mathbb{C}^{N\times d}, \quad \operatorname{Im}\psi_0 = 0. \tag{3}$$

The state starts real and the propagator makes it complex. A second network $V(C(X)) \in \mathbb{R}^{N\times d}$ produces a learned per-node potential that conditions the dynamics on node content throughout the evolution.

3.2 GROUPED SPECTRAL PROPAGATION

We split the d channels into H equal groups that we call the spectral heads. Head h carries a positive damping rate γ_h and a positive relative scale s_h together with real dispersion coefficients in

$$\omega_h(\lambda) = \omega_{1h}\lambda + \omega_{2h}\lambda^2. \tag{4}$$

The quadratic term is a higher-order correction to the linear Schrödinger dispersion. We parameterize γ_h and s_h with softplus and normalize the scales to

$$\frac{1}{H}\sum_{h=1}^{H} s_h = 1, \tag{5}$$

so the heads differ in relative speed and not in overall magnitude. Every layer advances the state by the same duration Δt so that T layers cover a total time $T\Delta t$. One layer first applies the spectral propagator to the channels of head h,

$$\widetilde{\psi}_{t,h} = U \exp\big[-\Delta t\, s_h\gamma_h\Lambda - i\Delta t\, s_h(\omega_{1h}\Lambda + \omega_{2h}\Lambda^2)\big]\, U^\top \psi_{t,h}. \tag{6}$$

A slow and heavily damped head acts like a smoothing filter, whereas a fast dispersive head keeps high-frequency structure and transports it. We evaluate all heads together in one operator application per layer. The split between decay and phase is exact because the damping sets the real part of the exponent and the dispersion sets its imaginary part. Let P_h denote the matrix function in Equation 6, let

$$g_h(\lambda) = \exp\big[-\Delta t\, s_h\gamma_h\lambda - i\Delta t\, s_h(\omega_{1h}\lambda + \omega_{2h}\lambda^2)\big] \tag{7}$$

be its scalar response, and let $\|\cdot\|_F$ denote the Frobenius norm throughout. All proofs are in Appendix B.

Theorem 1 (Spectral energy law)**.** *Let $L = U\Lambda U^\top$ with U real orthogonal and $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$, and let every λ_ℓ lie in $[0, 2]$. Let $P_h = g_h(L)$ be the matrix function that Equation 7 defines through functional calculus (Higham, 2008), and let $\Delta t, s_h, \gamma_h \geq 0$. For any $\psi_h \in \mathbb{C}^{N\times d/H}$, write $\widehat{\psi}_h = U^\top\psi_h$. Then*

$$\|P_h\psi_h\|_F^2 = \sum_{\ell=1}^{N} e^{-2\Delta t\, s_h\gamma_h\lambda_\ell}\|\widehat{\psi}_{h,\ell:}\|_2^2 \leq \|\psi_h\|_F^2. \tag{8}$$

Let $\mathcal{Q}_{\geq\lambda_0}$ be the orthogonal projector onto the Laplacian modes with eigenvalue at least λ_0. Then

$$\|\mathcal{Q}_{\geq\lambda_0}P_h\psi_h\|_F \leq e^{-\Delta t\, s_h\gamma_h\lambda_0}\|\mathcal{Q}_{\geq\lambda_0}\psi_h\|_F. \tag{9}$$

Finally, let $\widetilde{P}_h = p_h(L)$ be any polynomial realization of the same response and let

$$\varepsilon_h = \sup_{\lambda\in[0,2]} \big|p_h(\lambda) - g_h(\lambda)\big| \tag{10}$$

be its uniform error on the Laplacian spectrum. Then

$$\|(\widetilde{P}_h - P_h)\psi_h\|_F \leq \varepsilon_h\|\psi_h\|_F, \qquad \|\widetilde{P}_h\psi_h\|_F \leq (1+\varepsilon_h)\|\psi_h\|_F. \tag{11}$$

The identity shows that dispersion changes phase and preserves spectral energy: a head with zero damping preserves the norm however strongly it transports, while positive damping suppresses every band away from the Laplacian nullspace at an explicit rate. The two regimes the bank must hold at once therefore have separate parameters, and the last part of the theorem carries the guarantee over to the sparse realization below.

One learned complex channel map M that is shared across layers then mixes the channels before a deterministic local modulation closes the layer:

$$z_t = \widetilde{\psi}_t M, \tag{12}$$

$$q_{t+1} = z_t \odot \exp[-\Delta t\, s \odot \mu \odot |z_t|^2] \odot \exp[-i\Delta t\, s \odot V(C(X))], \tag{13}$$

$$\psi_{t+1} = \mathcal{C}_B(\mathrm{Re}\, q_{t+1}). \tag{14}$$

The potential V makes the phase velocity depend on node content while the positive absorption μ saturates extreme amplitudes. The vectors s, μ expand the per-head values across their channels whereas V is specific to a node and a channel. The cubic term enters the modulus and not the argument, so the nonlinearity this model carries is dissipative rather than Hamiltonian, which gives the layer an intrinsic amplitude ceiling: Proposition 6 caps every entry the modulation emits at a value no state can exceed, whereas at $\mu = 0$ the modulation becomes a pure phase rotation that only the clip below bounds. The layer closes by passing on the real part, which the entrywise radial map $\mathcal{C}_B$ clips at the fixed radius $B = 50$, and the encoded initial state carries the same safeguard. Alternating a linear spectral flow with a pointwise nonlinear map is the structure of split-step integrators for nonlinear dispersive equations (Trotter, 1959; Hairer et al., 2006). Up to the learned channel map, the real projection and the radial safeguard, one layer is a first-order split step of the damped graph nonlinear Schrödinger equation

$$i\,\partial_t \psi_h = s_h \left[\omega_h(L)\psi_h + V_h \odot \psi_h - i\left(\gamma_h L \psi_h + \mu_h |\psi_h|^2 \odot \psi_h\right)\right], \tag{15}$$

whose cubic term sits on the dissipative axis and so places it in the damped rather than the conservative branch of the nonlinear Schrödinger family (Sulem & Sulem, 2007). With linear dispersion and without the damping, the absorption and the potential, the equation reduces to the graph analogue of the free Schrödinger equation. Equations 6–14 define the complete layer.

A layer passes its real part to the next one while the propagator, the channel map and the modulation all stay complex, and projecting does not remove the dispersion. With the response written as $g_h(\lambda) = e^{-\Delta t\, s_h \gamma_h \lambda}(\cos[\Delta t\, s_h \omega_h(\lambda)] - i \sin[\Delta t\, s_h \omega_h(\lambda)])$, the real part carried forward is a channel-map combination of two terms that can each change sign in λ, while the modulation is untouched because it reads the state only through $|z_t|^2$. The projection removes a phase that would persist across layers rather than the oscillatory spectral response.

Proposition 2 (Depth profile of the response)**.** *Write $P_h(\tau)$ for the exact propagator of head h at duration τ. Then $P_h(\tau)P_h(\sigma) = P_h(\tau + \sigma)$ for all $\tau, \sigma \geq 0$, so for every head*

$$\prod_{t=1}^{T} g_h(\lambda) = \exp\left[-T\Delta t\, s_h\left(\gamma_h \lambda + i\omega_h(\lambda)\right)\right], \qquad \left|\prod_{t=1}^{T} g_h(\lambda)\right| = e^{-T\Delta t\, s_h \gamma_h \lambda}. \tag{16}$$

The composed modulus is at most one at every frequency and exactly one on the Laplacian nullspace, whatever the depth.

The family is closed under composition, so stacking layers cannot take the response out of it and no depth attenuates the lowest mode. A polynomial response of degree $K - 1$ composes to one of degree $T(K - 1)$ that obeys no such constraint at any frequency.

We evaluate Equation 6 in the eigenbasis whenever a dense eigendecomposition is affordable. On larger graphs a Chebyshev expansion of the propagator applies the same filter with sparse matrix products (Defferrard et al., 2016). Its polynomial order follows the magnitude of the exponent, and this keeps the uniform error ε_h of Equation 10 small across the parameter range the search explores. Proposition 5 bounds that error explicitly and shows it to decay faster than any geometric rate in the order (Trefethen, 2019). The coefficients are recomputed from the current spectral parameters at every call so that gradients reach them, and the recurrence advances the real and imaginary parts together in one sparse product against the symmetric operator. One layer costs $O(K|\mathcal{E}|d + Nd^2)$ operations at polynomial order K, and recurrence activations are recomputed in the backward pass whenever storing them would exceed a fixed memory budget.

3.3 Readout and objective

The readout turns the final state into decoder features and we select one of two variants per dataset. The intensity variant uses $|\psi_T|^2$. The signed variant concatenates $\mathrm{Re}\,\psi_T$ beside it and so keeps the sign that squaring discards. A layer boundary passes on a real state, so these two quantities are all the readout can form and neither reads anything beyond the evolved state. Two linear decoders map those features to class logits and we read every metric from their mean.

The state reaching the readout obeys the following bound.

Theorem 3 (Full-depth state bound)**.** *Let $m = \|M\|_2$ be the spectral norm of the channel map, with $m = 1$ when mixing is absent, and let $\varepsilon = \max_h \varepsilon_h$ with $\varepsilon = 0$ for exact propagation. Assume*

$\mu_h \geq 0$ *for every head. Then*

$$\|\psi_T\|_F \leq \min\Big\{B\sqrt{Nd}, \big[m(1+\varepsilon)\big]^T \|\psi_0\|_F\Big\}. \tag{17}$$

For exact propagation the bound is independent of the dispersion and the potential, so the only potentially expansive learned factor before the radial safeguard is the channel map.

We call the linear spectral flow contractive in operator norm and the nonlinear modulation norm-nonincreasing, and neither claim says that either map is a pairwise contraction. Depth and intensity coupling can therefore reshape the state without inflating it, while in the ablations of Section 4.3 we test whether the single expansive factor the bound admits contributes to accuracy.

Under label scarcity a detector needs a calibrated decision boundary and an accurate ranking of its scores, and the objective trains one decoder for each. The first decoder trains with focal modulation,

$$\mathcal{L}_{\text{cls}}(\hat{y}) = \frac{1}{|\mathcal{T}|} \sum_{v \in \mathcal{T}} (1 - p_{v,y_v})^{\eta} \, \text{CE}(p_v, y_v), \tag{18}$$

where $\mathcal{T}$ is the training set and η is the focal exponent. We apply focal reweighting per example rather than per class so that no explicit class prior enters the loss. The cross entropy includes label smoothing (Szegedy et al., 2016). This loss calibrates a decision boundary. However, the ranking metrics that practitioners use depend on the order of the whole score list, which a threshold loss tracks only loosely. The second decoder therefore learns the ordering directly. Write s_v for its anomaly score and σ for the logistic function. The smooth total and positive ranks of a positive anchor i are

$$r_i = \sum_{j \in \mathcal{T}} \sigma\left(\frac{s_j - s_i}{\zeta}\right), \qquad r_i^+ = \sum_{j \in \mathcal{T}:\, y_j = 1} \sigma\left(\frac{s_j - s_i}{\zeta}\right), \tag{19}$$

with temperature ζ. The ranking loss over the positive training nodes $\mathcal{T}_+$ is

$$\mathcal{L}_{\text{AP}} = 1 - \frac{1}{|\mathcal{T}_+|} \sum_{i \in \mathcal{T}_+} \frac{r_i^+}{r_i}, \tag{20}$$

a differentiable surrogate of one minus average precision. A label-smoothed cross entropy CE keeps the ranking decoder calibrated, and the total loss is

$$\mathcal{L} = \mathcal{L}_{\text{cls}}(\hat{y}^{\text{cls}}) + \mathcal{L}_{\text{AP}}(\hat{y}^{\text{rk}}) + \beta \, \text{CE}(\hat{y}^{\text{rk}}, y), \tag{21}$$

and the score is $\hat{y} = \frac{1}{2}(\hat{y}^{\text{cls}} + \hat{y}^{\text{rk}})$. The temperature $\zeta = 0.1$ and the weight $\beta = 0.2$ are constants of the method rather than searched values, so no dataset receives an objective selected for it. The surrogate scales as the positive count times the training-set size, so we process anchors in chunks to bound peak memory. We train with Adam (Kingma & Ba, 2015) and keep an exponential moving average of the weights for evaluation (Tarvainen & Valpola, 2017) in the spirit of Polyak–Ruppert averaging (Polyak & Juditsky, 1992). That average stabilizes the epoch selection of Section 4.1.

No part of the model refers to node identity, from the deterministic context through to the decoders.

Proposition 4 (Permutation equivariance)**.** *Let $F_\theta(A, X)$ be the evaluation-mode node logits of G-SNLS. For every node-permutation matrix Π,*

$$F_\theta(\Pi A \Pi^\top, \Pi X) = \Pi F_\theta(A, X). \tag{22}$$

This holds for the exact eigenspace operator as well as its Chebyshev realization, and it covers the context encoder, the potential and the nonlinear evolution together with both readout variants and both decoders.

4 Experiments

4.1 Dataset & Experimental Settings

Datasets. We evaluate on eight established benchmarks carrying 10K to 204K nodes and anomaly rates from 2.98% to 21.82%: organic fraud on Amazon, T-Finance and Elliptic, platform-behavior

Table 1: Performance evaluation in terms of test AUROC over ten independent trials. Bold marks the best mean and underlining the runner-up, with an asterisk on a lead the paired test resolves.

Dataset	GCN	GraphSAGE	GPR-GNN	GRAND	BWGNN	AMNet	GHRN	SAGAD	EGNN	G-SNLS
Tolokers	79.51±0.75	82.78±0.67	76.99±0.78	73.33±0.80	82.58±0.82	78.52±0.62	82.60±0.83	82.13±0.90	81.34±0.68	**85.11±0.75***
Questions	74.89±1.37	73.90±1.04	72.82±1.16	71.13±1.08	73.08±0.77	71.19±1.23	72.97±0.74	76.51±1.00	71.20±1.38	**76.78±1.27**
Amazon	88.47±0.74	97.83±0.29	96.95±0.37	96.24±0.66	97.62±0.33	97.43±0.55	97.58±0.41	97.10±0.57	96.49±0.89	**98.01±0.32**
T-Finance	95.82±0.43	96.70±0.25	94.98±0.74	92.18±0.45	96.67±0.30	95.80±0.46	96.60±0.29	94.47±0.49	95.08±0.57	**97.01±0.33***
Elliptic	98.01±0.14	98.51±0.20	98.89±0.09	95.09±0.14	98.47±0.11	98.49±0.10	98.44±0.14	98.37±0.13	98.64±0.13	**99.23±0.12***
inj-Amazon	76.07±1.36	84.55±1.66	89.19±1.11	82.33±1.54	86.23±1.47	83.81±2.48	86.01±1.65	**91.41±0.90**	89.60±1.97	90.18±2.09
Minesweeper	73.50±2.31	91.64±0.55	89.50±0.44	71.30±0.71	91.01±0.44	75.30±0.89	90.79±0.53	90.65±0.50	91.05±0.57	**95.55±0.56***
inj-Flickr	84.19±0.84	82.03±1.93	91.83±0.21	77.02±0.71	93.17±0.55	87.42±3.12	93.52±0.35	94.81±0.21	89.59±0.64	**95.08±0.14***

Table 2: Performance evaluation in terms of test AUPRC over ten independent trials. Bold marks the best mean and underlining the runner-up, with an asterisk on a lead the paired test resolves.

Dataset	GCN	GraphSAGE	GPR-GNN	GRAND	BWGNN	AMNet	GHRN	SAGAD	EGNN	G-SNLS
Tolokers	49.74±1.50	53.60±1.80	44.55±1.74	37.19±1.77	53.36±1.45	45.50±1.83	53.30±1.68	52.48±1.57	51.97±1.79	**59.66±1.76***
Questions	13.44±1.01	16.97±2.57	17.59±1.79	13.39±1.35	17.45±1.60	16.35±1.73	17.50±1.38	21.02±2.00	11.39±1.53	**23.02±2.43***
Amazon	50.21±2.37	91.25±1.12	87.89±1.61	86.33±1.94	90.16±1.29	90.21±1.64	90.06±1.51	89.37±1.15	87.90±2.64	**91.75±1.02***
T-Finance	80.94±1.61	84.87±0.56	81.19±1.54	74.00±1.40	85.25±1.06	82.21±1.42	85.30±1.21	76.56±1.94	81.05±1.33	**86.40±0.88***
Elliptic	91.11±0.74	94.37±0.56	94.54±0.46	70.37±1.00	93.47±0.65	93.71±0.60	93.47±0.51	93.48±0.58	94.44±0.65	**96.82±0.36***
inj-Amazon	43.54±2.75	44.03±6.34	37.73±2.52	21.27±1.80	37.65±3.78	23.60±4.44	35.69±3.93	64.57±2.82	61.89±3.67	**70.74±9.16**
Minesweeper	39.61±3.02	73.22±1.51	66.44±0.75	36.55±1.19	71.03±1.24	41.10±1.43	71.15±1.31	69.55±0.86	70.61±2.01	**85.84±1.82***
inj-Flickr	54.33±1.54	20.61±0.92	37.39±0.67	15.61±0.42	63.00±1.34	23.68±7.34	62.25±2.51	67.77±1.00	61.46±2.23	**75.13±0.72***

labels on Tolokers, Questions and Minesweeper, and injected outliers on inj-Amazon and inj-Flickr. Camouflage makes anomalies smooth on the homophilous fraud graphs whereas they are rough on the heterophilous or nearly featureless ones, so no single spectral response suffices across the suite.

Protocol. We divide labeled nodes into ten independent splits with 40% training, 20% validation and 40% test nodes. We train every method once per split with Adam for at most three hundred epochs and with a patience of one hundred on validation AUROC, and we read both ranking metrics at the epoch of best validation AUROC under one selection rule for every method. We read macro F1 at its best validation epoch and threshold (Appendix D). We take hyperparameters from a two-stage search that screens on the first split and rescreens on the first two before freezing the choice for all ten. Those development splits are the first two evaluation splits rather than a separate pool, and the search reads validation labels only, so no test node influences any selection.

Baselines. We compare against GCN (Kipf & Welling, 2017), GraphSAGE (Hamilton et al., 2017), GPR-GNN (Chien et al., 2021), GRAND (Chamberlain et al., 2021), BWGNN (Tang et al., 2022), AMNet (Chai et al., 2022), GHRN (Gao et al., 2023), SAGAD (Liu et al., 2026b) and EGNN (Liu et al., 2026a). We rerun every one of them on the same splits under the same search procedure. AMNet weights a fixed pair of spectral channels per node, so the comparison with it measures what learned dissipative dynamics add over such a pair.

4.2 Experimental Results

We compare G-SNLS with every baseline on the two ranking metrics in Tables 1 and 2. Bold and a full shade mark the best mean while underlining and a light shade mark the runner-up. An asterisk marks a lead that survives a paired Wilcoxon test after Holm correction across the nine baselines of the same dataset and metric. G-SNLS leads the strongest baseline on fifteen of sixteen comparisons, and that test resolves twelve of them. No baseline holds a resolved lead over G-SNLS on any dataset.

Average Rank. We report the average rank of each method over the eight datasets in Table 4. G-SNLS takes the leading rank on both metrics, and the Friedman test rejects rank equality among the ten methods on each. Every method that the Nemenyi test separates from G-SNLS commits to a response fixed before the graph arrives, whether that is the low-pass filter of GCN, the diffusion equation of GRAND, the signed polynomial of GPR-GNN or the Bernstein pair AMNet reweights per node. Those that stay in the group of G-SNLS carry more than one band, edit the graph or keep a self term that the propagation cannot smooth away. With eight datasets the critical difference is 4.79, so we compare G-SNLS with that group through the per-dataset tests.

AUROC. The propagator bank and its intensity coupling account for the AUROC gains, which appear where one frequency response cannot serve every node. Minesweeper represents one extreme because

Table 3: Performance evaluation in terms of test macro F1 over ten independent trials. Bold marks the best mean and underlining the runner-up, with an asterisk on a lead the paired test resolves.

Dataset	GCN	GraphSAGE	GPR-GNN	GRAND	BWGNN	AMNet	GHRN	SAGAD	EGNN	G-SNLS
Tolokers	66.94±0.81	69.95±1.47	63.34±1.21	60.91±0.82	69.95±0.93	65.87±0.85	69.81±0.87	69.85±1.17	69.10±0.49	**72.53±1.13***
Questions	59.14±0.59	60.12±1.11	60.27±0.50	58.37±0.55	60.11±0.78	59.60±0.76	59.89±1.25	62.22±0.79	56.73±0.99	**62.26±0.98**
Amazon	73.52±0.94	**92.32±0.39**	90.34±1.00	89.77±1.01	91.77±0.58	91.82±0.49	91.46±0.92	90.78±0.44	91.32±0.46	91.92±0.86
T-Finance	87.72±1.24	89.61±0.38	87.84±1.05	85.46±1.03	90.18±0.71	89.03±0.99	**90.39±0.56**	85.85±0.77	88.50±0.79	90.31±0.44
Elliptic	92.07±0.29	94.62±0.43	94.05±0.25	82.44±0.53	93.94±0.34	94.15±0.19	94.13±0.37	94.02±0.26	94.42±0.50	**95.77±0.15***
inj-Amazon	77.98±0.84	76.39±1.97	74.37±1.59	62.74±1.07	70.09±1.56	61.58±2.72	71.12±2.22	83.82±1.02	82.06±1.29	**84.97±1.05***
Minesweeper	62.68±1.66	78.69±0.89	75.86±0.45	61.98±0.73	77.27±0.73	65.05±0.72	76.86±0.57	76.70±0.72	77.66±0.66	**85.08±1.35***
inj-Flickr	80.00±0.38	61.73±0.34	73.93±0.41	59.09±0.38	79.24±0.57	65.82±4.96	79.26±0.49	80.94±0.32	77.20±1.13	**83.21±0.43***

Table 4: Average rank over the eight datasets for both ranking metrics, lower being better. A shared group letter means the Nemenyi test cannot separate the ranks.

	GCN	GraphSAGE	GPR-GNN	GRAND	BWGNN	AMNet	GHRN	SAGAD	EGNN	**G-SNLS**
Average AUROC rank	7.62	4.00	6.12	9.75	4.25	7.00	4.75	4.88	5.50	**1.12**
Nemenyi group	bc	ab	bc	c	ab	bc	ab	ab	abc	**a**
Average AUPRC rank	7.75	4.12	6.00	9.75	4.62	6.62	4.62	4.75	5.75	**1.00**
Nemenyi group	bc	ab	bc	c	ab	bc	ab	ab	abc	**a**

Table 5: Change in test AUROC over ten independent trials when one component is removed.

Variant	Minesweeper	Tolokers	Questions	Amazon	Elliptic	inj-Flickr
Full model AUROC (%)	95.55±0.56	85.08±0.76	76.87±1.21	98.01±0.32	99.24±0.12	95.10±0.15
No nonlinear coupling	−3.71**	−2.22**	−7.95**	−0.21*	−1.11**	−3.96**
No nonlinear absorption	−0.66**	−1.43**	−5.23**	−0.09*	−0.07**	−1.66**
No channel mixing	−1.94**	−6.63**	−5.66**	−0.16**	−0.18**	−6.72**
No spectral damping	−0.23	−1.49**	−4.97**	−0.15*	−0.27**	−0.65**
No dispersion	−3.86**	−1.12**	+0.12	+0.02	−0.03	−0.66**
Single frequency head	−0.62**	−0.37**	–	+0.04	−0.07	–
No input context	+0.65	–	−0.49	−0.51**	–	–
Intensity readout	−0.28	–	−0.60	+0.04	−0.06*	+0.14*

its mine signal resides at high frequency. Deleting dispersion is the largest single-component loss on that graph while deleting damping changes nothing the paired test resolves. Tolokers represents the other extreme. It is a dense network of benign collaboration where damped heads suppress the background, and there the removal of damping causes a resolved loss instead. No fixed response satisfies both extremes, whereas the per-head parameterization accommodates both. AMNet performs comparably on the homophilous fraud graphs and trails by 20.24% on Minesweeper, which indicates that a pair of bands fixed in advance suffices where anomalies are smooth but not where the response must change shape. On inj-Amazon the fused two-band response of SAGAD already recovers the planted outliers, and the paired test does not separate it from G-SNLS.

AUPRC. Every margin widens on average precision. AUROC averages over the whole score list that the negatives dominate, whereas average precision weights the head of that list where positives are scarce and the background sets the ceiling. The dissipative half of the dynamics controls that head, since absorption and the learned potential suppress amplitude that the intensity coupling would otherwise spread into normal nodes. This effect is largest on inj-Flickr, where the margin of G-SNLS over the strongest baseline grows from a fraction of a percent on AUROC to over 7% on average precision. Deleting the learned potential there is the largest single-component effect of the whole study. On inj-Amazon the margin changes sign and G-SNLS leads on average precision.

Macro F1. We report the threshold-based comparison in Table 3. G-SNLS attains the best mean on six of the eight datasets at an average rank of 1.25. The paired test resolves five of those six leads and separates no leader from G-SNLS on Amazon or T-Finance. The widest margins again fall on the two extremes of the AUROC comparison, at 6.39% on Minesweeper and 2.59% on Tolokers.

Table 6: Free-response control in terms of test AUROC over ten independent trials, where K is the matched polynomial order and Δ the change against the per-split scores of Table 1.

		Schrödinger	Free, matched order		Free, order 32		Free, random init	
Dataset	K	AUROC (%)	AUROC (%)	Δ (%)	AUROC (%)	Δ (%)	AUROC (%)	Δ (%)
Minesweeper	10	95.55±0.56	95.21±0.49	−0.33	93.01±0.57	-2.54^{**}	93.86±0.52	-1.68^{**}
Tolokers	15	85.11±0.75	84.52±0.74	-0.58^{**}	83.47±0.93	-1.64^{**}	80.67±0.90	-4.43^{**}
Questions	15	76.78±1.27	77.07±0.92	+0.28	77.21±0.98	$+0.43^{*}$	71.85±2.52	-4.93^{**}
Amazon	18	98.01±0.32	97.96±0.33	−0.05	97.88±0.34	−0.12	97.34±0.40	-0.67^{**}
Elliptic	18	99.23±0.12	99.24±0.13	+0.01	99.18±0.08	−0.05	98.76±0.15	-0.47^{**}
inj-Flickr	15	95.08±0.14	95.18±0.27	+0.10	91.35±5.77	-3.73^{*}	93.75±1.08	-1.33^{**}

4.3 Importance of Each Component

We toggle one component at a time on six datasets that span the suite (Table 5) and retrain every variant on the same ten splits under the validated configuration of its dataset. A paired Wilcoxon test judges the change in mean test AUROC against the rerun full model. One asterisk marks $p < 0.05$ and two mark $p < 0.01$, and shading repeats those marks in red for a loss and teal for a gain. A dash marks a toggle the validated configuration already applies. Removing either of two components causes a resolved loss on every graph. Deleting the nonlinear modulation costs between 0.21% and 7.95% AUROC because it leaves a learned linear filter bank behind, and deleting the complex channel map costs between 0.16% and 6.72% because the heads then evolve in isolation. Deleting only the intensity-dependent absorption in the modulation also causes a resolved loss on all six graphs at a cost of 0.07% to 5.23%. By Proposition 6 the same absorption caps every entry the modulation emits, so the term that bounds the amplitude also raises accuracy. The two linear terms of the propagator divide the datasets instead because damping is resolved on five of the six graphs and unresolved on Minesweeper while dispersion is the largest single loss on that same grid and neutral on Questions, where the validated dynamics reduce to one damped head. Questions shows the widest spread, with resolved losses when we remove the intensity coupling, the channel map or the damping. Changes on Amazon and Elliptic stay small because both are close to the performance ceiling.

4.4 Constrained Versus Free Response

We replace the response of every head by an unconstrained complex Chebyshev polynomial on the six ablation datasets and keep the rest of the validated model fixed (Appendix G). This free filter carries K complex coefficients per head and layer at the order the adaptive rule of Section 3.2 selects, and it starts from the Chebyshev interpolant of the Schrödinger response so that both arms begin from the same operator. We report the test AUROC of every arm in Table 6 and mark resolved changes as in Section 4.3. At matched order the free filter changes AUROC by a mean of -0.09%, and a Wilcoxon test across the six datasets returns $p = 0.69$. Constraining each head to four parameters shared across layers therefore preserves accuracy. At order 32 the free filter collapses on three of the ten inj-Flickr splits. By Theorem 1 and Proposition 2 the constrained response amplifies no band at any width or depth and so admits no such failure. From random coefficients the free filter loses between 0.47% and 4.93% on all six datasets. The constrained parameterization also gives the training loss the lowest peak curvature per unit change in the response of any arm on all six datasets (Appendix H).

5 Conclusion

G-SNLS replaces the single spectral response of existing graph anomaly detectors with a bank of energy-controlled Schrödinger propagators whose damping and dispersion are learned per head and tied to node content by an intensity coupling between steps. We separate energy-preserving dispersion from spectral decay, show that composition keeps the response in its family, bound the state at full depth and prove equivariance. The method takes the leading rank on both ranking metrics against baselines selected by the same search. Against a free response of matched degree, the constraint preserves accuracy while it gives a response that no width or depth can make diverge and a loss surface whose most curved direction is flatter per unit change in the response on every graph we measure. Spectral sketches for web-scale and temporal graphs remain open.

AI USE STATEMENT

We used generative AI tools to assist with implementation of the method and evaluation pipeline, refinement of hypotheses and their framing, design of the experimental protocol, and polishing the presentation and formulation of mathematical claims. All mathematical claims and AI-assisted work were reviewed and verified by the authors. The authors take responsibility for the final content of the submission.

REPRODUCIBILITY

All datasets are public and are loaded through standard libraries. The appendices specify the data processing, the model configurations, the training procedures, the baseline settings, the software and hardware, and complete proofs of the theoretical claims. Everything needed to reproduce the reported results is in the paper and in the anonymized source code submitted as supplementary material.

ETHICS STATEMENT

This work collects no new data and recruits no human subjects. The benchmarks are public graphs with anonymized features. The supplementary code reproduces the reported evaluation and gives no capability for identifying individuals in a live system.

REFERENCES

Guoguo Ai, Hezhe Qiao, Hui Yan, and Guansong Pang. Semi-supervised graph anomaly detection via robust homophily learning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Deyu Bo, Xiao Wang, Chuan Shi, and Huawei Shen. Beyond low-frequency information in graph convolutional networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 3950–3957, 2021.

Ziwei Chai, Siqi You, Yang Yang, Shiliang Pu, Jiarong Xu, Haoyang Cai, and Weihao Jiang. Can abnormality be detected by graph neural networks? In *IJCAI*, pp. 1945–1951, 2022.

Ben Chamberlain, James Rowbottom, Maria I Gorinova, Michael Bronstein, Stefan Webb, and Emanuele Rossi. Grand: Graph neural diffusion. In *International conference on machine learning*, pp. 1407–1418. PMLR, 2021.

Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized pagerank graph neural network. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=n6jl7fLxrP.

Jeongwhan Choi, Seoyoung Hong, Noseong Park, and Sung-Bae Cho. Gread: Graph neural reaction-diffusion networks. In *International conference on machine learning*, pp. 5722–5747. PMLR, 2023.

Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.

Kaize Ding, Jundong Li, Rohit Bhanushali, and Huan Liu. Deep anomaly detection on attributed networks. In *Proceedings of the 2019 SIAM international conference on data mining*, pp. 594–602. SIAM, 2019.

Yingtong Dou, Zhiwei Liu, Li Sun, Yutong Deng, Hao Peng, and Philip S Yu. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pp. 315–324, 2020.

Moshe Eliasof, Eldad Haber, and Eran Treister. Pde-gcn: Novel architectures for graph neural networks motivated by partial differential equations. *Advances in neural information processing systems*, 34:3836–3849, 2021.

Stefano Fiorini, Stefano Coniglio, Michele Ciavotta, and Enza Messina. Sigmanet: One laplacian to rule them all. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 7568–7576, 2023.

Yuan Gao, Xiang Wang, Xiangnan He, Zhenguang Liu, Huamin Feng, and Yongdong Zhang. Addressing heterophily in graph anomaly detection: A perspective of graph spectrum. In *Proceedings of the ACM web conference 2023*, pp. 1528–1538, 2023.

Ernst Hairer, Christian Lubich, and Gerhard Wanner. *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, volume 31. Springer-Verlag Berlin, 2006.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Mingguo He, Zhewei Wei, Zengfeng Huang, and Hongteng Xu. Bernnet: Learning arbitrary graph spectral filters via bernstein approximation. *Advances in neural information processing systems*, 34:14239–14251, 2021.

Nicholas J Higham. *Functions of matrices: theory and computation*. SIAM, 2008.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017. URL https://openreview.net/forum?id=SJU4ayYgl.

Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Kay Liu, Yingtong Dou, Yue Zhao, Xueying Ding, Xiyang Hu, Ruitong Zhang, Kaize Ding, Canyu Chen, Hao Peng, Kai Shu, et al. Bond: Benchmarking unsupervised outlier node detection on static attributed graphs. *Advances in neural information processing systems*, 35:27021–27035, 2022.

Yang Liu, Xiang Ao, Zidi Qin, Jianfeng Chi, Jinghua Feng, Hao Yang, and Qing He. Pick and choose: a gnn-based imbalanced learning approach for fraud detection. In *Proceedings of the web conference 2021*, pp. 3168–3177, 2021.

Yilin Liu, Hongchao Zhang, Ahmad Taha, Taylor T Johnson, and Meiyi Ma. Modeling spectral energy shifts in spatio-temporal graph anomaly detection. In *Forty-third International Conference on Machine Learning*, 2026a. URL https://openreview.net/forum?id=sgNhWbqXfJ.

Yunhui Liu, Qizhuo Xie, Yinfeng Chen, Xudong Jin, Tao Zheng, Bin Chong, and Tieke He. Mitigating homophily disparity in graph anomaly detection: A scalable and adaptive approach. In *Proceedings of the ACM Web Conference 2026*, pp. 1127–1138, 2026b.

Julian John McAuley and Jure Leskovec. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In *Proceedings of the 22nd international conference on World Wide Web*, pp. 897–908, 2013.

Hoang NT and Takanori Maehara. Revisiting graph neural networks: All we have is low-pass filters. *arXiv preprint arXiv:1905.09550*, 2019.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1ldO2EFPr.

Oleg Platonov, Denis Kuznedelev, Michael Diskin, Artem Babenko, and Liudmila Prokhorenkova. A critical look at the evaluation of GNNs under heterophily: Are we really making progress? In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=tJbbQfw-5wv.

Boris T Polyak and Anatoli B Juditsky. Acceleration of stochastic approximation by averaging. *SIAM journal on control and optimization*, 30(4):838–855, 1992.

T Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. Graph-coupled oscillator networks. In *International conference on machine learning*, pp. 18888–18909. PMLR, 2022.

Catherine Sulem and Pierre-Louis Sulem. *The nonlinear Schrödinger equation: self-focusing and wave collapse*. Springer Science & Business Media, 2007.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016.

Jianheng Tang, Jiajin Li, Ziqi Gao, and Jia Li. Rethinking graph neural networks for anomaly detection. In *International conference on machine learning*, pp. 21076–21089. PMLR, 2022.

Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30, 2017.

Lloyd N Trefethen. *Approximation theory and approximation practice, extended edition*. SIAM, 2019.

Hale F Trotter. On the product of semi-groups of operators. *Proceedings of the American Mathematical Society*, 10(4):545–551, 1959.

Yuelin Wang, Kai Yi, Xinliang Liu, Yu Guang Wang, and Shi Jin. ACMP: Allen-cahn message passing with attractive and repulsive forces for graph neural networks. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=4fZc_79Lrqs.

Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I Weidele, Claudio Bellei, Tom Robinson, and Charles E Leiserson. Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv preprint arXiv:1908.02591*, 2019.

Felix Wu, Amauri Souza, Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Weinberger. Simplifying graph convolutional networks. In *International conference on machine learning*, pp. 6861–6871. Pmlr, 2019.

Xitong Zhang, Yixuan He, Nathan Brugnone, Michael Perlmutter, and Matthew Hirn. Magnet: A neural network for directed graphs. *Advances in neural information processing systems*, 34: 27003–27015, 2021.

Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. *Advances in neural information processing systems*, 33:7793–7804, 2020.

APPENDIX

A ALGORITHM SUMMARY

We collect the complete pipeline of Section 3 into executable steps in Algorithm 1. Lines 1 and 2 run once per graph and are reused across every epoch and trial. The remaining lines are one training step, and the loop over t is the layer of Equations 6–14. Dropout on the encoder input and on the readout features is left out for readability.

Algorithm 1 G-SNLS forward pass and training step

Require: Graph with normalized adjacency $\widehat{A}$ and Laplacian L, features X, depth T, heads H, step Δt, clip radius B
Ensure: Evaluation-mode node logits

$(U, \Lambda) \leftarrow \mathrm{eigh}(L)$ if affordable, else prepare the Chebyshev recurrence on L ▷ once per graph
$C \leftarrow$ deterministic context of X ▷ validated variant, Section 3.1
repeat
  $s \leftarrow \mathrm{softplus}(\tilde{s})$ rescaled to unit mean
  $\gamma \leftarrow \mathrm{softplus}(\tilde{\gamma})$, $\mu \leftarrow \mathrm{softplus}(\tilde{\mu})$, $\omega_h(\lambda) \leftarrow \omega_{1h}\lambda + \omega_{2h}\lambda^2$
  $\psi_0 \leftarrow \mathcal{C}_B(E(C))$, $V \leftarrow V(C)$
  for $t = 1$ to T **do**
    $\widetilde{\psi}_{t,h} \leftarrow U \exp[-\Delta t\, s_h(\gamma_h \Lambda + i\omega_h(\Lambda))] U^\top \psi_{t-1,h}$ for all h ▷ Eq. 6
    $z_t \leftarrow \widetilde{\psi}_t M$ ▷ complex channel map
    $\psi_t \leftarrow \mathcal{C}_B\big(\Re\left[z_t \odot \exp[-\Delta t\, s \odot \mu \odot |z_t|^2 - i\Delta t\, s \odot V]\right]\big)$ ▷ Eq. 14
  end for
  $\Phi \leftarrow \mathrm{readout}(\psi_T)$ ▷ validated variant, Section 3.3
  $\hat{y}^{\mathrm{cls}} \leftarrow D_{\mathrm{cls}}(\Phi)$, $\hat{y}^{\mathrm{rk}} \leftarrow D_{\mathrm{rk}}(\Phi)$
  $\mathcal{L} \leftarrow \mathcal{L}_{\mathrm{cls}}(\hat{y}^{\mathrm{cls}}) + \mathcal{L}_{\mathrm{AP}}(\hat{y}^{\mathrm{rk}}) + \beta\, \mathrm{CE}(\hat{y}^{\mathrm{rk}}, y)$ ▷ Eq. 21
  $\theta \leftarrow \mathrm{Adam}(\theta, \nabla_\theta \mathcal{L})$, $\bar{\theta} \leftarrow \alpha\bar{\theta} + (1-\alpha)\theta$
until the validation metric stops improving
return $\frac{1}{2}(\hat{y}^{\mathrm{cls}} + \hat{y}^{\mathrm{rk}})$ evaluated at the averaged weights $\bar{\theta}$

B PROOFS OF THEORETICAL RESULTS

B.1 NOTATION AND CONVENTIONS

Throughout this appendix $L = I - \widehat{A}$ is the symmetric normalized Laplacian of an undirected graph, so L is real symmetric and positive semidefinite and its spectrum satisfies $\mathrm{spec}(L) \subseteq [0, 2]$ (Chung, 1997). The spectral theorem supplies an eigendecomposition $L = U\Lambda U^\top$ with U real orthogonal. For a scalar function f defined on $\mathrm{spec}(L)$ we write

$$f(L) = U f(\Lambda) U^\top, \qquad f(\Lambda) = \mathrm{diag}\big(f(\lambda_1), \ldots, f(\lambda_N)\big), \tag{23}$$

a definition that does not depend on the eigenbasis chosen for a repeated eigenvalue (Higham, 2008). Two standard norm facts are used without further comment (Horn & Johnson, 2012). The Frobenius norm is invariant under an orthogonal or unitary factor,

$$\|UZ\|_F = \|Z\|_F \qquad \text{for every } Z \in \mathbb{C}^{N \times k}, \tag{24}$$

and it is compatible with the spectral norm on either side,

$$\|MZ\|_F \le \|M\|_2 \|Z\|_F, \qquad \|ZM'\|_F \le \|Z\|_F \|M'\|_2. \tag{25}$$

We state every result in evaluation mode, where dropout is disabled and the averaged weights are used.

B.2 Spectral energy law

Proof of Theorem 1. By Equation 23 the propagator is $P_h = U g_h(\Lambda) U^\top$. Applying Equation 24 to the leading factor U and writing $\widehat{\psi}_h = U^\top \psi_h$, we obtain

$$\|P_h \psi_h\|_F^2 = \|g_h(\Lambda)\widehat{\psi}_h\|_F^2 = \sum_{\ell=1}^{N} |g_h(\lambda_\ell)|^2 \|\widehat{\psi}_{h,\ell:}\|_2^2, \tag{26}$$

where the second equality expands the Frobenius norm row by row and $\widehat{\psi}_{h,\ell:}$ is the ℓ-th row. The exponent of Equation 7 splits into the real part $-\Delta t\, s_h \gamma_h \lambda$ and the purely imaginary part $-i\Delta t\, s_h \omega_h(\lambda)$, and $|e^z| = e^{\operatorname{Re} z}$, so

$$|g_h(\lambda)| = e^{-\Delta t\, s_h \gamma_h \lambda} \le 1 \qquad \text{for every } \lambda \in [0, 2], \tag{27}$$

the inequality holding because $\Delta t, s_h, \gamma_h \ge 0$ and $\lambda \ge 0$. Substituting Equation 27 into Equation 26 proves Equation 8.

The projector $\mathcal{Q}_{\ge\lambda_0} = U \operatorname{diag}(\mathbf{1}[\lambda_\ell \ge \lambda_0]) U^\top$ is itself a function of L, so it commutes with P_h, and the two are diagonal in the same basis. Restricting the sum in Equation 26 to the indices with $\lambda_\ell \ge \lambda_0$ and bounding each surviving multiplier by $e^{-\Delta t\, s_h \gamma_h \lambda_0}$ proves Equation 9. That bound is legitimate because $\lambda \mapsto e^{-\Delta t\, s_h \gamma_h \lambda}$ is nonincreasing.

For the polynomial realization, functional calculus applied to $p_h - g_h$ gives

$$\widetilde{P}_h - P_h = U \operatorname{diag}\big(p_h(\lambda_\ell) - g_h(\lambda_\ell)\big)_{\ell=1}^{N} U^\top, \tag{28}$$

whose spectral norm $\max_\ell |p_h(\lambda_\ell) - g_h(\lambda_\ell)|$ is at most ε_h because $\operatorname{spec}(L) \subseteq [0, 2]$ while Equation 10 takes the supremum over that interval. The first bound of Equation 11 then follows from Equation 25. For the second, the triangle inequality and $\|P_h\|_2 = \max_\ell |g_h(\lambda_\ell)| \le 1$ give

$$\|\widetilde{P}_h \psi_h\|_F \le \|P_h \psi_h\|_F + \|(\widetilde{P}_h - P_h)\psi_h\|_F \le (1 + \varepsilon_h)\|\psi_h\|_F. \tag{29}$$

□

B.3 Chebyshev interpolation error

Theorem 1 leaves ε_h unquantified. The next proposition bounds it for the realization the implementation uses and justifies the claim of Section 3.2 that a modest order suffices over the parameter range the search explores.

Proposition 5 (Chebyshev interpolation error)**.** *Fix a layer and a head with $\gamma_h \ge 0$ and write $a = \Delta t\, s_h \ge 0$. Let p_h be the degree-$(K-1)$ Chebyshev interpolant of g_h at the K Chebyshev nodes of $[0, 2]$, which is the realization the implementation computes. For $\varrho > 1$ put*

$$R_\varrho = 1 + \tfrac{1}{2}\left(\varrho + \varrho^{-1}\right). \tag{30}$$

Then the uniform error of Equation 10 obeys

$$\varepsilon_h \le \frac{4\varrho^{-(K-1)}}{\varrho - 1} \exp\!\left[a\left((\gamma_h + |\omega_{1h}|)R_\varrho + |\omega_{2h}|R_\varrho^2\right)\right] \qquad \textit{for every } \varrho > 1. \tag{31}$$

Proof. The affine change of variable $\lambda = 1 + x$ maps $[-1, 1]$ onto $[0, 2]$, and $x \mapsto g_h(1 + x)$ is entire because g_h is the exponential of a polynomial. Let E_ϱ be the Bernstein ellipse of parameter $\varrho > 1$, which is the image of the circle $|z| = \varrho$ under $x = \frac{1}{2}(z + z^{-1})$. Every $x \in E_\varrho$ satisfies $|x| \le \frac{1}{2}(\varrho + \varrho^{-1})$, so the corresponding $\lambda = 1 + x$ satisfies $|\lambda| \le R_\varrho$. Using $|e^z| = e^{\operatorname{Re} z} \le e^{|z|}$ and the triangle inequality on Equation 4, we obtain

$$|g_h(\lambda)| \le \exp\!\left[a\left(\gamma_h|\lambda| + |\omega_{1h}||\lambda| + |\omega_{2h}||\lambda|^2\right)\right] \le \exp\!\left[a\left((\gamma_h + |\omega_{1h}|)R_\varrho + |\omega_{2h}|R_\varrho^2\right)\right] =: M_\varrho \tag{32}$$

on that ellipse. The classical convergence theory for a function analytic inside E_ϱ and bounded there by M_ϱ (Trefethen, 2019, Ch. 8) bounds the uniform error of its degree-$(K-1)$ Chebyshev interpolant on $[-1, 1]$ by $4M_\varrho\varrho^{-(K-1)}/(\varrho - 1)$ and bounds that of the corresponding truncation by half as much. Undoing the change of variable gives Equation 31. □

Error Decay in the Order. The bound of Equation 31 holds for arbitrarily large ϱ because g_h is entire, and its exponent grows only linearly in ϱ when $\omega_{2h} = 0$ and quadratically otherwise. Reading the bound at $\varrho = K$ in the first case and at $\varrho = \sqrt{K}$ in the second gives

$$\varepsilon_h \le 4e^{2ac_1}\Big(\frac{e^{ac_1/2}}{K}\Big)^{K-1} \quad (K \ge 2), \qquad \varepsilon_h \le 4e^{a(2c_1+4c_2)}\Big(\frac{e^{a(c_1/4+c_2)}}{\sqrt{K}}\Big)^{K-1} \quad (K \ge 4), \tag{33}$$

with $c_1 = \gamma_h + |\omega_{1h}|$ and $c_2 = |\omega_{2h}|$, so ε_h decays faster than any fixed geometric rate in K.

Adaptive Order Rule. The adaptive rule of Section 3.2 raises K with the magnitude of the exponent that Equation 31 makes the error depend on. Once $\omega_{2h} = 0$ the response is the pure exponential $e^{c\lambda}$ with $c = -a(\gamma_h + i\omega_{1h})$, and its coefficients in the basis $T_k(\lambda - 1)$ are $e^c I_0(c)$ for $k = 0$ and $2e^c I_k(c)$ for $k \ge 1$ with I_k the modified Bessel function of the first kind. The bound $|I_k(c)| \le (|c|/2)^k e^{|c|^2/4}/k!$ shrinks by the factor $|c|/(2k+2)$ from each order to the next and so by more than half once k exceeds $|c|$, and below its ceiling of 64 the rule sets K above $3|c|$.

B.4 Depth profile of the response

Proof of Proposition 2. The exact response of head h at duration τ is

$$g_h(\lambda;\tau) = e^{-\tau c_h(\lambda)}, \qquad c_h(\lambda) = s_h\big[\gamma_h\lambda + i\omega_h(\lambda)\big]. \tag{34}$$

The scalar identity $e^{-\tau c}e^{-\sigma c} = e^{-(\tau+\sigma)c}$ gives $g_h(\lambda;\tau)g_h(\lambda;\sigma) = g_h(\lambda;\tau+\sigma)$ at every $\lambda \in [0,2]$. All operators $P_h(\tau) = g_h(L;\tau)$ are functions of the same symmetric matrix, so by Equation 23 one U diagonalizes them all and they commute while their products multiply symbols. The semigroup property therefore extends from the symbol to the operator, and composing T layers of duration Δt gives $e^{-T\Delta t\, c_h(\lambda)}$. Its modulus is the exponential of the real part of the exponent, and $\operatorname{Re} c_h(\lambda) = s_h\gamma_h\lambda$ because ω_h is real. That quantity is nonnegative for $\lambda \in [0,2]$ and $s_h, \gamma_h \ge 0$ so that the modulus $e^{-T\Delta t\, s_h\gamma_h\lambda}$ is at most one. It vanishes at $\lambda = 0$ and the modulus is one there at every depth. □

B.5 Full-depth state bound

Proof of Theorem 3. Let $h(c)$ be the head containing channel c and let $\mathcal{N}_t$ denote the nonlinear modulation of Equation 14 before the radial safeguard. Its two exponential factors are a real decay and a unit-modulus phase that carries the potential, so entrywise

$$\big|[\mathcal{N}_t(z)]_{vc}\big| = |z_{vc}|\exp\big[-\Delta t\, s_{h(c)}\mu_{h(c)}|z_{vc}|^2\big] \le |z_{vc}|, \tag{35}$$

using $\Delta t, s_h \ge 0$ and $\mu_h \ge 0$. Summing squared moduli gives

$$\|\mathcal{N}_t(z)\|_F \le \|z\|_F. \tag{36}$$

The projection onto the real part satisfies $|\operatorname{Re} z_{vc}| \le |z_{vc}|$ entrywise and is therefore norm-nonincreasing, and the radial map $\mathcal{C}_B$ that follows it either leaves an entry alone or rescales it to modulus B below its own and is norm-nonincreasing for the same reason.

For the linear half of the layer, Theorem 1 applied to the block of each head and summed over heads gives

$$\|\widetilde{P}\psi\|_F^2 = \sum_{h=1}^{H}\|\widetilde{P}_h\psi_h\|_F^2 \le (1+\varepsilon)^2\sum_{h=1}^{H}\|\psi_h\|_F^2 = (1+\varepsilon)^2\|\psi\|_F^2, \tag{37}$$

with $\varepsilon = \max_h \varepsilon_h$. The channel map multiplies from the right, so the second inequality of Equation 25 gives $\|\psi M\|_F \le m\|\psi\|_F$. Composing propagation, mixing, modulation, the real projection and the clip in the order the layer applies them gives

$$\|\psi_t\|_F \le m(1+\varepsilon)\|\psi_{t-1}\|_F, \tag{38}$$

and iterating from ψ_0 proves the power arm of Equation 17. For the other arm take $T \ge 1$. The last layer closes with $\mathcal{C}_B$, whose output has entries of modulus at most B. Every entry of ψ_T therefore has modulus at most B, so $\|\psi_T\|_F \le B\sqrt{Nd}$. At $T = 0$ the same bound holds because the encoded initial state carries the same safeguard. □

B.6 ABSORPTION CEILING

The bound of Theorem 3 comes from the radial safeguard, so it is a property of the clip and not of the dynamics. The next proposition shows that a positive absorption supplies a ceiling of its own that holds whatever the state does.

Proposition 6 (Absorption caps the state amplitude). *Fix a layer t and a head h and write $a = \Delta t\, s_h > 0$. Let $\mathcal{N}_t$ be the modulation of Equation 14, whose phase on entry (v, c) of that head is $-aV_{vc}$, and write A for the smallest ceiling on the modulus of an entry leaving layer t that holds for every input z of the modulation.*

1. *If $\mu_h = 0$ then $|[\mathcal{N}_t(z)]_{vc}| = |z_{vc}|$ for every entry, so the modulation is a pure phase rotation that passes amplitude through untouched, and the radial safeguard is the only operation that bounds it. Then $A = B$.*

2. *If $\mu_h > 0$ then every entry of the modulated state obeys*

$$\big|[\mathcal{N}_t(z)]_{vc}\big| \le \frac{1}{\sqrt{2e\, a\mu_h}}, \tag{39}$$

with equality at $|z_{vc}|^2 = 1/(2a\mu_h)$. That bound is a property of the modulation alone and holds whatever the state does, while the safeguard that closes the layer contributes B, so

$$A = \min\big\{B,\ (2e\, a\mu_h)^{-1/2}\big\}. \tag{40}$$

The modulation therefore binds strictly before the safeguard exactly when $2e\, a\mu_h B^2 > 1$, and in that regime the two ceilings stand in the ratio $(2e\, a\mu_h B^2)^{-1/2} < 1$.

Proof. Both parts use one construction for minimality. Setting $z_{vc} = re^{iaV_{vc}}$ for $r \ge 0$ cancels the phase the modulation applies, so q_{vc} is real and nonnegative. The projection that follows leaves it alone and the radial map then returns $\min\{q_{vc}, B\}$.

For part 1 the choice $\mu_h = 0$ makes the real exponential factor of Equation 14 equal to one, so Equation 35 holds with equality and the modulation preserves every modulus. The remaining operations are the real projection, which only decreases a modulus, and the radial map, which caps one at B, so $A \le B$. Taking $r \ge B$ in the construction above emits an entry of modulus exactly B, so no smaller state-independent ceiling is available and $A = B$.

For part 2 write $r = |z_{vc}| \ge 0$. Equation 35 makes the modulated modulus $f(r) = re^{-a\mu_h r^2}$, and

$$f'(r) = e^{-a\mu_h r^2}\big(1 - 2a\mu_h r^2\big) \tag{41}$$

vanishes on $r > 0$ only at $r_\star = (2a\mu_h)^{-1/2}$, where it changes sign from positive to negative, so $r_\star$ is the global maximizer. Evaluating gives $f(r_\star) = (2a\mu_h)^{-1/2}e^{-1/2}$, which is Equation 39. Neither the real projection nor the radial map increases a modulus, so this bound survives to the end of the layer, and the radial map caps the same entry at B. Hence $A \le \min\{B, (2e\, a\mu_h)^{-1/2}\}$. The choice $r = r_\star$ in the construction above attains both cases of the minimum. It emits $q_{vc} = (2e\, a\mu_h)^{-1/2}$, which the radial map leaves untouched when it lies below B and returns as exactly B otherwise, so the entry leaves the layer at modulus $\min\{B, (2e\, a\mu_h)^{-1/2}\}$ and Equation 40 holds with equality. □

Binding Ceiling. The ceiling in part 1 is a property of the safeguard so that only a state reaching the clip radius attains it, whereas Equation 39 is a property of the modulation alone and binds whatever the state does. The parameters rather than the state determine which of them governs Equation 40. At every validated initialization $2e\, a\mu_h B^2$ lies between 502 and 15,073 across all heads, so the modulation binds first and the safeguard sits between 22 and 123 times above the amplitude a layer actually emits. Deleting μ_h therefore leaves that amplitude limited by a clip rather than by the dynamics, and we measure the resulting cost in the extended ablation of Appendix F.

B.7 PERMUTATION EQUIVARIANCE

Proof of Proposition 4. Let Π be an $N \times N$ permutation matrix and write $A^{\Pi} = \Pi A \Pi^{\top}$ and $X^{\Pi} = \Pi X$. Relabeling permutes the degrees and the isolated nodes so that $D^{\Pi} = \Pi D \Pi^{\top}$ and $I_0^{\Pi} = \Pi I_0 \Pi^{\top}$, and since Π is orthogonal and both are diagonal,

$$\widehat{A}^{\Pi} = (D^{\Pi})^{-1/2} A^{\Pi} (D^{\Pi})^{-1/2} + I_0^{\Pi} = \Pi \widehat{A} \Pi^{\top}, \qquad L^{\Pi} = I - \widehat{A}^{\Pi} = \Pi L \Pi^{\top}. \tag{42}$$

Every term of the deterministic context is formed from products with $\widehat{A}$ and from entrywise operations while Π permutes rows without mixing them, so

$$C_{A^{\Pi}}(X^{\Pi}) = \Pi\, C_A(X) \tag{43}$$

term by term. Equation 42 gives $\widehat{A}^{\Pi} X^{\Pi} = \Pi \widehat{A} X$ and $(\widehat{A}^{\Pi})^2 X^{\Pi} = \Pi \widehat{A}^2 X$, while absolute values, products, clamping at the positive floor and square roots all act entrywise and commute with a row permutation. The encoder E and the potential network V are shared across nodes, so each applies the same map to every row and preserves Equation 43.

For any scalar function f defined on $\mathrm{spec}(L)$, Equations 23 and 42 give

$$f(L^{\Pi}) = f(\Pi L \Pi^{\top}) = \Pi f(L) \Pi^{\top}, \tag{44}$$

which holds for any admissible eigenbasis because functional calculus does not depend on the one selected (Higham, 2008). Hence $f(L^{\Pi})\Pi\psi = \Pi f(L)\psi$, and this covers the exact propagator together with the Chebyshev recurrence, which evaluates the polynomial f as a matrix polynomial in L and so never needs an eigenbasis. Channel mixing multiplies from the right and so acts on channels, whereas the nonlinear modulation and the radial map act entrywise. All three commute with a row permutation, and induction over the T layers therefore gives

$$\psi_T^{\Pi} = \Pi \psi_T, \tag{45}$$

and the real projection at the layer boundary is entrywise and does not disturb it.

Both readout variants are formed from entrywise moduli and concatenation along channels, so they commute with Π while both decoders are row-wise. Composing the steps proves Equation 22. Dropout is disabled in evaluation mode so that no node-dependent randomness breaks the chain. □

C DATASETS

Organic Fraud. Amazon connects reviewers of the musical-instruments category who reviewed the same products, gave the same star rating or wrote similar text, and helpful-vote ratios identify the fraudulent reviewers (McAuley & Leskovec, 2013; Dou et al., 2020). We exclude nodes outside the reliably labeled block from every split under the standard convention. T-Finance connects anonymized accounts through transaction records while annotators mark the fraud, money-laundering and gambling accounts (Tang et al., 2022). Elliptic is a Bitcoin transaction graph whose labeled part separates licit from illicit transactions (Weber et al., 2019). Its unlabeled transactions stay in the graph for propagation but never enter a split.

Heterophilous Graphs. Tolokers connects crowdsourcing workers who shared a task, and the banned workers are the positives (Platonov et al., 2023). Questions connects users of a question-answering service when one answered the other within a one-year window. We take the minority of users who did not remain active as positives so that the rare class is the detection target. Minesweeper is a synthetic grid graph where each cell links to its adjacent cells, and one fifth of the cells are mines (Platonov et al., 2023). We include these three graphs because their near-zero adjusted homophily or near-absent features stress frequency adaptivity.

Injected Outliers. The remaining two benchmarks plant contextual and structural outliers into a co-purchase graph and a photo-sharing network under the established injection protocol (Ding et al., 2019; Liu et al., 2022), and any injected node of inj-Amazon or inj-Flickr counts as an anomaly.

Graph Statistics. Table 7 lists the scale, feature dimension and anomaly rate of every benchmark after these loading and labeling conventions are applied. The counts therefore refer to the graphs the models actually receive. We exclude self-loops from the edge count. The loaded Questions graph is

Table 7: Dataset statistics as loaded by the evaluation pipeline, giving node, distinct undirected edge and feature counts with the labeled nodes eligible for splits and their anomaly rate.

Dataset	Nodes	Edges	Features	Labeled	Anomalies (%)
Tolokers	11,758	519,000	10	11,758	21.82
Questions	48,921	153,540	301	48,921	2.98
Amazon	11,944	4,398,392	25	8,639	9.50
Minesweeper	10,000	39,402	7	10,000	20.00
T-Finance	39,357	21,222,543	10	39,357	4.58
Elliptic	203,769	234,355	165	46,564	9.76
inj-Amazon	13,752	257,435	767	13,752	5.05
inj-Flickr	89,250	470,311	500	89,250	4.95

the only one that carries them, one at every node. The only graph with isolated nodes is inj-Amazon. Equation 1 places all 276 of them in the Laplacian nullspace, where the aggregation leaves each unchanged and every spectral response acts on such a node through its value at $\lambda = 0$. Elliptic shows 165 features because the standard loader drops the raw time step that Weber et al. (2019) count among their 166.

Random Splits. We resample random node splits per trial on every dataset under the standard convention in supervised anomaly detection, and that resampling makes the ten trials independent. Elliptic is the one graph whose nodes carry a time stamp, so a random split lets a model see transactions that come after the ones it is scored on. We keep the shared convention instead of special-casing one dataset, and every method is affected the same way.

D Macro F1 results

Threshold Selection. We report the threshold-based comparison against every baseline over the same ten splits of Section 4.1 in Table 3. Bold and a full shade mark the best mean while underlining and a light shade mark the runner-up. This metric needs one selection rule beyond the ranking protocol. We read macro F1 at the epoch that maximizes validation macro F1 and choose the decision threshold on validation data before carrying it unchanged to the test set for every method alike. Threshold-based evaluation rewards calibrated decision boundaries, and the objective ablation of Appendix F shows that the focal modulation this model carries for that purpose is worth at most 0.19% of ranking accuracy, so a lead on this metric reflects the representation rather than the loss.

Comparison Against Baselines. G-SNLS attains the best mean macro F1 on six of eight datasets and an average rank of 1.25, and the Friedman test rejects rank equality among the ten methods at $p < 0.001$. Five of the six leads are significant. The largest gains again come from frequency adaptivity, at 6.39% on Minesweeper and 2.59% on Tolokers. The Elliptic and inj-Flickr improvements of 1.16% and 2.27% indicate that the ranking-aware representation also improves the decision boundary. G-SNLS ranks second on the two densest graphs Amazon and T-Finance, where the paired test does not separate it from the leader.

E Component ablations

We report the architecture toggles of Section 4.3 together with the focal-loss toggle in Table 8 and separate the dispersion and the two terms of the modulation in Appendix F. A toggle that coincides with the validated configuration of a dataset is not run and appears as a dash. Figure 2 shows the Questions column of Table 5 in absolute terms.

Coupling and Channel Mixing. Removing either of two components causes a resolved loss on every ablation dataset. Removing the nonlinear modulation deletes the learned potential and the intensity-dependent absorption in one step, and it costs between 0.21% and 7.95%. That edit leaves a learned linear filter bank behind, so the gains come from the intensity-coupled terms of Equation 15 and not from the linear residue of the architecture. We remove the potential and the absorption of the modulation together in this table and separate them in Appendix F. Removing the complex channel map has a comparable cost of 0.16% to 6.72%. The map alone moves information between spectral

Table 8: Component ablations as the change in test AUROC over ten independent trials against the rerun full model, whose absolute AUROC heads the table.

Variant	Minesweeper	Tolokers	Questions	Amazon	Elliptic	inj-Flickr
Full model AUROC (%)	95.55±0.56	85.08±0.76	76.87±1.21	98.01±0.32	99.24±0.12	95.10±0.15
Architecture, Δ AUROC (%)						
No nonlinear coupling	-3.71^{**}	-2.22^{**}	-7.95^{**}	-0.21^{*}	-1.11^{**}	-3.96^{**}
No spectral damping	-0.23	-1.49^{**}	-4.97^{**}	-0.15^{*}	-0.27^{**}	-0.65^{**}
No channel mixing	-1.94^{**}	-6.63^{**}	-5.66^{**}	-0.16^{**}	-0.18^{**}	-6.72^{**}
Single frequency head	-0.62^{**}	-0.37^{**}	–	$+0.04$	-0.07	–
No input context	$+0.65$	–	-0.49	-0.51^{**}	–	–
Intensity readout	-0.28	–	-0.60	$+0.04$	-0.06^{*}	$+0.14^{*}$
Objective, Δ AUROC (%)						
No focal modulation	-0.04	–	-0.11	$+0.06$	-0.19^{**}	–

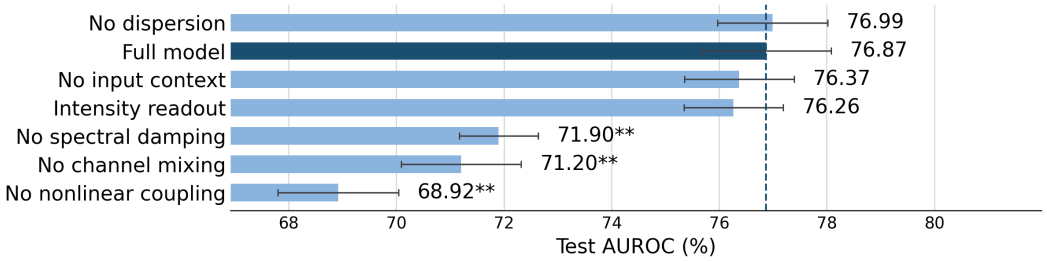

Figure 2: The component study on Questions in absolute terms. The dashed line marks the rerun full model while asterisks mark the paired Wilcoxon outcome against it, and error bars give one standard deviation over the ten trials. Bars are ordered by mean AUROC and carry the same arms the Questions column of Table 5 reports as changes.

scales, so without it the grouped propagators evolve as independent filters that only the decoder combines. It is also the only potentially expansive learned factor in Theorem 3, so its ablation tests whether that factor contributes to accuracy. Four of the six ablation datasets run several heads, and collapsing the bank to one head is resolved on two of them at 0.62% and 0.37%.

Damping and Dispersion. Damping is worth 4.97% on Questions and 1.49% on Tolokers, where suppressing high-frequency content exposes the anomalies. On Minesweeper the paired test resolves nothing because the mine signal on that grid lives at high frequency, and attenuation works against what the decoder needs. Dispersion shows the opposite pattern. It costs 3.86% on Minesweeper and 1.12% on Tolokers while it stays neutral on Questions, whose validated dynamics reduce to one damped head with nothing to transport. Learning decay and phase velocity per head lets the same architecture serve both patterns.

Context and Readout. The input context and the intensity readout sit outside the dynamics and affect the model far less than the propagator and its coupling do.

F Extended ablations

We break the two composite components of Appendix E into their parts in Table 9 on the same ten splits and under the same paired protocol against the rerun full model. The dynamics arms separate the dispersion of Equation 6 from the two terms of the nonlinear modulation while the objective arms separate the terms of the training loss. A dash again marks a toggle the validated configuration already applies.

Nonlinear Absorption. Deleting the intensity-dependent absorption is resolved on all six graphs and costs 5.23% on Questions and 1.66% on inj-Flickr. By Proposition 6 the absorption caps the amplitude a layer emits for every input whereas without it only the radial safeguard bounds that amplitude.

Table 9: Decomposed dynamics and objective ablations as the change in test AUROC over ten independent trials against the rerun full model, whose absolute AUROC heads the table.

Variant	Minesweeper	Tolokers	Questions	Amazon	Elliptic	inj-Flickr
Full model AUROC (%)	95.55±0.56	85.08±0.76	76.87±1.21	98.01±0.32	99.24±0.12	95.10±0.15
Dynamics decomposition, Δ AUROC (%)						
No dispersion	-3.86^{**}	-1.12^{**}	$+0.12$	$+0.02$	-0.03	-0.66^{**}
No nonlinear absorption	-0.66^{**}	-1.43^{**}	-5.23^{**}	-0.09^{*}	-0.07^{**}	-1.66^{**}
No learned potential	-0.66^{*}	-0.04	-1.16^{**}	$+0.01$	$+0.07^{*}$	-15.71^{**}
Objective decomposition, Δ AUROC (%)						
No focal modulation	-0.04	–	-0.11	$+0.06$	-0.19^{**}	–
No ranking decoder	$+0.20$	-0.24	-0.41^{*}	-0.05	$+0.05$	$+0.16^{*}$

Learned Potential. The effect of the learned potential is far more uneven. It costs 15.71% on inj-Flickr as the largest single-component effect in this study, against 1.16% on Questions. The validated configuration accounts for this spread. The inj-Flickr configuration runs one head on raw features, so the potential is the only term that conditions the flow on node content once the encoder has acted, and deleting the potential leaves a node-independent propagator.

Partial Versus Complete Removal. Deleting the potential alone costs more on inj-Flickr than deleting the whole modulation it belongs to, which Table 5 puts at 3.96%. The two arms leave different models behind rather than nested ones. Removing the modulation removes the absorption with the potential and hands the next layer a linear filter bank, whereas removing the potential alone keeps an intensity-dependent absorption that now damps every node by the same rule. On a graph whose only source of node conditioning was the potential, that surviving term suppresses the amplitude differences the decoder reads, so the partial edit is worse than the complete one.

Dispersion. Deleting the dispersion reduces every propagator to a monotone spectral decay. It costs 3.86% on Minesweeper as the largest linear effect there and 1.12% on Tolokers, and it stays neutral on Questions, whose validated configuration pairs one head with strong damping. That cost arises from the sign-changing frequency response itself and not from any phase the state carries between layers, since a layer boundary carries only the real part.

Objective. The objective arms change the metric far less than the dynamics arms do. Deleting the focal modulation produces changes between a 0.19% loss and a 0.06% gain. Removing the ranking decoder costs 0.41% on Questions, whose positives are scarcest and whose average precision the decoder was added to protect. Neither loss term accounts for the gains of Section 4.2.

G Free-response control

Every arm of Appendices E and F deletes a term from the Schrödinger response, so each one compares that response against a smaller member of its own family. In this appendix we compare it against a larger member. ChebNet, GPR-GNN and BernNet learn polynomial responses (Defferrard et al., 2016; Chien et al., 2021; He et al., 2021), and a free polynomial of the same degree contains the Chebyshev realization of the Schrödinger response as a special case. Substituting the free polynomial therefore isolates the exact contribution of the constraint of Theorem 1 and Proposition 2.

The arms below test whether the constrained response adds accuracy at matched degree, what property of the initial response matters among free filters of matched degree and whether a free filter beyond matched degree fails for lack of a norm bound or for lack of ordinary regularization.

G.1 Control design

Unconstrained Response. We replace the response of each head by an unconstrained Chebyshev polynomial in the Laplacian,

$$g_h(\lambda) = \sum_{k=0}^{K-1} c_{t,h,k} T_k(\lambda - 1), \qquad c_{t,h,k} \in \mathbb{C}, \tag{46}$$

Table 10: Initial response of the free filter at matched order, as the change in test AUROC against the per-split scores of Table 1 over ten independent trials.

	Schrödinger	Δ test AUROC (%) by initial response					
Dataset	AUROC (%)	Schrödinger	Monotone decay	Beta wavelet	PPR	Random	Random, bounded
Minesweeper	95.55±0.56	−0.33	−1.06**	+0.20	−1.17**	−1.68**	−0.96**
Tolokers	85.11±0.75	−0.58**	−0.76**	−4.46**	−0.87**	−4.43**	−8.90**
Questions	76.78±1.27	+0.28	+0.29	−6.86**	−0.20	−4.93**	−5.90**
Amazon	98.01±0.32	−0.05	−0.04	−0.07	−0.02	−0.67**	−0.49**
Elliptic	99.23±0.12	+0.01	+0.04	−0.07	+0.05*	−0.47**	−0.48**
inj-Flickr	95.08±0.14	+0.10	+0.14*	−0.83**	+0.10	−1.33**	−0.42*

Table 11: The order-32 free filter under coefficient decay and under the supremum bound of Theorem 1, read as in Table 10. A trial fails when it falls more than 5% AUROC below its reference.

	Order 32		+ coefficient decay			+ bounded response	
Dataset	Δ (%)	Failed	λ	Δ (%)	Failed	Δ (%)	Failed
Minesweeper	−2.54**	0/10	10^{-1}	−2.23**	0/10	−2.67**	0/10
Tolokers	−1.64**	0/10	10^{-3}	−1.62**	0/10	−1.37**	0/10
Questions	+0.43*	0/10	10^{-2}	+0.41	0/10	+0.06	0/10
Amazon	−0.12	0/10	10^{-1}	−0.06	0/10	−0.10	0/10
Elliptic	−0.05	0/10	10^{-1}	−0.03	0/10	−0.01	0/10
inj-Flickr	−3.73*	3/10	10^{-1}	+0.01	0/10	+0.05	0/10

and change nothing else. The deterministic context, complex encoder and channel map all stay at the validated setting of each dataset, as do the nonlinear modulation, the readout, the decoder and the objective. Six datasets are covered across both operator realizations of Section 3.2. Amazon is the one benchmark whose graph is dense enough for the eigenbasis form while the other five use the Chebyshev recurrence.

Paired Reference. Every arm is paired against the per-split test AUROC of Table 1 rather than a rerun full model, a choice that removes run-to-run drift from the comparison and separates this control from the ablations. The two references agree exactly on Minesweeper and Amazon and to at most 0.09% of mean AUROC on the other four datasets. Every table of this appendix marks a resolved change with one asterisk at $p < 0.05$ and two at $p < 0.01$, and shading repeats those marks.

Choices Favoring the Free Filter. Three choices favor the free filter. Its coefficients are free per layer as well as per head, so it subsumes the head scales that the propagator no longer reads while the proposed model keeps four spectral parameters per head shared across all layers. Those coefficients are also initialized to the Chebyshev interpolant of the Schrödinger response at the validated initial values and consume no random draw. The two arms therefore begin from the same operator to within 10^{-5} at every order we run, and they share their encoder and decoder weights. The order K reported in Table 6 is the one the adaptive rule of Section 3.2 selects for the Schrödinger response itself, so the free filter never receives less resolution than the response it replaces. We also run $K = 32$, far beyond what any validated configuration needs.

Initialization Arms. We add two families of further arms that vary what the free filter is given. The first varies what it starts from. In place of the Schrödinger interpolant we initialize the coefficients from the monotone decay that remains once the dispersion is removed, from the beta-wavelet bands of BWGNN (Tang et al., 2022), from the personalized-PageRank response of GPR-GNN (Chien et al., 2021) or from noise. The wavelet arm gives one band to each head, and the PageRank arm one teleport probability from the grid its authors report, so the bank starts as diverse as the Schrödinger response makes it. Every one of these responses is rescaled to the unit supremum on $[0, 2]$ that the Schrödinger exponential already attains at the Laplacian nullspace, so the arms differ in the shape of the initial response and not in its gain. The noise arm is the exception. It is left at the scale $K^{-1/2}$ at which the expected squared modulus of the response is at most one on $[0, 2]$ and exactly one at the Laplacian nullspace, and Appendix H reports its realized supremum.

Constraint Arms. The second family varies what constrains the filter once training starts. One arm imposes the conclusion of Theorem 1 on the free response directly. The coefficients of each layer and head are rescaled at every step so that $\sup_{\lambda\in[0,2]} |g_h(\lambda)| \leq 1$ on a grid fine enough to resolve a degree-31 polynomial, which is the property the Schrödinger response has by construction and the free filter lacks. Another gives the coefficients a weight decay of their own that validation selects from $\{10^{-3}, 10^{-2}, 10^{-1}\}$, since the decay suiting the rest of the network was never chosen for thirty-two free complex coefficients per head and layer. Both apply to the order-32 filter, and the bound is also run on the noise start at matched order.

Missing Guarantees. Equation 46 has no counterpart to Theorem 1 because an unconstrained coefficient vector can amplify any band and no frequency is forced to contract. The product bound of Theorem 3 then survives only through the radial safeguard. The free response also has no counterpart to Proposition 2. The Schrödinger response contracts everywhere and preserves $\lambda = 0$ exactly, so composing T layers neither grows the state nor empties it. A free response can fail in either direction. Leaving the low band above one makes depth amplify while pushing the whole response below one makes depth annihilate.

G.2 Findings

Equivalence at Matched Degree. The free response matches the accuracy of the Schrödinger response on this suite. Its per-dataset changes are -0.33, -0.58, $+0.28$, -0.05, $+0.01$ and $+0.10$ for a mean of -0.09 and a median of -0.02, and a Wilcoxon test over the six datasets returns $p = 0.69$. The one resolved per-dataset change favors the Schrödinger response. We also search the polynomial order and the learning rate on Minesweeper, Tolokers and Questions. Validation prefers the matched order to order 32 on all three, and the selected arms land at -0.33, -0.65 and $+0.31$ against -0.33, -0.58 and $+0.28$.

Divergence Beyond Matched Degree. Given thirty-two coefficients where ten to eighteen suffice, the free filter loses a mean of 1.28% AUROC, and two different mechanisms underlie that mean. Minesweeper and Tolokers shift uniformly downward by 2.54% and 1.64% with every split affected and none failing. On inj-Flickr the filter instead collapses on exactly three splits in ten while the remaining seven stay within 0.14% of the reference on average, so a mean loss of 3.73% comes with a standard deviation of 5.77%. A paired rank test is structurally insensitive to a failure that affects only a minority of the splits, so we report the collapse count beside the mean in Table 11 rather than in place of it. The collapsing splits differ across datasets and arms, so the failure is stochastic rather than tied to particular splits. No arm at matched degree collapses on any split of any dataset. Theorem 1 excludes this behavior by construction, since the Schrödinger response has no coefficient that can amplify a band and admits no analogue of this failure at any depth or width.

Separation of Failure Modes by Remedy. We give the widened filter either a coefficient weight decay or the supremum bound and report the outcome in Table 11. Both remedies remove the stochastic collapse. On inj-Flickr the order-32 filter loses 3.73% and fails three splits in ten, while a weight decay of 10^{-1} on the coefficients returns it to $+0.01\%$ and the supremum bound of Theorem 1 to $+0.05\%$, each with no split failing. Neither remedy corrects the uniform downward shift. The decay recovers 0.31% of the 2.54% the widened filter loses on Minesweeper and 0.02% of the 1.64% it loses on Tolokers, and imposing the bound recovers 0.27% on Tolokers while costing a further 0.13% on Minesweeper. A projection or a decay strong enough to hold the coefficients down therefore reproduces the part of the constraint that keeps a band from being amplified. No arm of this control recovers the uniform loss of 2.54% on Minesweeper and 1.64% on Tolokers.

Value of the Schrödinger Initialization. We replace the Schrödinger interpolant by coefficients drawn at scale $K^{-1/2}$, which bounds the expected squared modulus of the response by one with equality at the Laplacian nullspace and so matches the scale of what it replaces. That costs between 0.47% and 4.93% AUROC significantly on every one of the six datasets, and it collapses splits on Tolokers and Questions. This is one of only two arms of the control that resolve across datasets rather than within them, the other being the same start under the supremum bound. It loses on six of six at a mean of 2.25%, and a Wilcoxon test over the datasets returns $p = 0.03$ against $p = 0.69$ for the matched-order arm. On Questions the free filter of the same degree and capacity gains 0.28% when it starts from the Schrödinger response and loses 4.93% when it starts from noise. The parameterization therefore contributes a norm-controlled coordinate system with four parameters per head in place of

Table 12: Geometry of the response parameterization under the spectral measure of each graph, with conditioning read as the ratio of the largest retained singular value to the smallest.

Quantity	Parameterization	Minesweeper	Tolokers	Questions	Amazon	Elliptic	inj-Flickr
Identifiable directions	Schrödinger	12 of 16	12 of 16	3 of 4	12 of 16	12 of 16	2 of 3
	Free, Schrödinger init	576 of 640	896 of 960	104 of 120	320 of 576	576 of 576	224 of 240
	Free, monotone decay init	576 of 640	896 of 960	104 of 120	320 of 576	576 of 576	224 of 240
	Free, beta-wavelet init	576 of 640	896 of 960	104 of 120	320 of 576	576 of 576	224 of 240
	Free, random init	576 of 640	896 of 960	104 of 120	320 of 576	576 of 576	224 of 240
	Free, order 32	1600 of 2048	1792 of 2048	200 of 256	448 of 1024	1024 of 1024	448 of 512
	Free, order 32, bounded	1504 of 2048	1761 of 2048	197 of 256	440 of 1024	1008 of 1024	430 of 512
Conditioning	Schrödinger	57.2 / 96.4	51.4 / 64.4	6.3 / 6.24	116 / 117	12.3 / 13	1.14 / 1.15
	Free, Schrödinger init	186 / 186	82.9 / 82.9	373 / 373	364 / 364	4.77 / 4.77	158 / 158
	Free, monotone decay init	186 / 186	82.9 / 82.9	373 / 373	364 / 364	4.77 / 4.77	158 / 158
	Free, beta-wavelet init	186 / 186	82.9 / 82.9	373 / 373	364 / 364	4.77 / 4.77	158 / 158
	Free, random init	186 / 186	82.9 / 82.9	373 / 373	364 / 364	4.77 / 4.77	158 / 158
	Free, order 32	729 / 729	352 / 352	237 / 237	487 / 487	8.55 / 8.55	885 / 885
	Free, order 32, bounded	223 / 399	788 / 990	286 / 395	793 / 768	22.3 / 29.1	967 / 996
Peak gain over depth	Schrödinger	1 / 1	1 / 1	1 / 1	1 / 1	1 / 1	1 / 1
	Free, Schrödinger init	1 / 0.879	1 / 8.08	1 / 1.37	1 / 1.55	1 / 7.29	1 / 2.38
	Free, monotone decay init	1 / 0.808	1 / 11.5	1 / 1.29	1 / 1.55	1 / 6.88	1 / 8.76
	Free, beta-wavelet init	1 / 5	1 / 52.3	1 / 1.13	1 / 3.03	1 / 3.69	1 / 1.29
	Free, random init	1.01 / 1.33	6.26 / 21.2	0.602 / 0.742	1.07 / 2.35	1.5 / 5.9	0.188 / 31.8
	Free, order 32	1 / 2.72	1 / 20	1 / 1.44	1 / 1.52	1 / 7.98	1 / 403
	Free, order 32, bounded	1 / 0.172	1 / 0.501	1 / 0.947	1 / 0.904	1 / 0.948	1 / 0.5
Median gain over depth	Schrödinger	0.387 / 0.538	0.161 / 0.235	0.00167 / 0.00209	0.168 / 0.179	0.161 / 0.171	2.8e-06 / 4.85e-06
	Free, Schrödinger init	0.387 / 0.27	0.161 / 0.333	0.00167 / 0.00347	0.168 / 0.229	0.161 / 0.164	2.8e-06 / 1.95e-06
	Free, monotone decay init	0.387 / 0.242	0.161 / 0.302	0.00167 / 0.00235	0.168 / 0.228	0.161 / 0.174	2.8e-06 / 2.52e-06
	Free, beta-wavelet init	0.00013 / 0.00141	0.00013 / 0.0028	0.311 / 0.404	0.0114 / 0.0133	0.0114 / 0.0378	0.097 / 0.349
	Free, random init	0.0165 / 0.0337	0.00625 / 0.0176	0.115 / 0.126	0.111 / 0.123	0.0667 / 0.0745	0.00557 / 0.00821
	Free, order 32	0.387 / 0.362	0.161 / 0.255	0.00167 / 0.00232	0.168 / 0.243	0.161 / 0.159	2.8e-06 / 7.29e-05
	Free, order 32, bounded	0.387 / 0.0088	0.161 / 0.0421	0.00167 / 0.00228	0.168 / 0.15	0.161 / 0.0532	2.8e-06 / 1.07e-09
Response supremum	Schrödinger	1.000 / 1.000	1.000 / 1.000	1.000 / 1.000	1.000 / 1.000	1.000 / 1.000	1.000 / 1.000
	Free, Schrödinger init	1.000 / 1.388	1.000 / 1.816	1.000 / 1.256	1.000 / 1.275	1.000 / 2.037	1.000 / 1.889
	Free, monotone decay init	1.000 / 1.508	1.000 / 1.845	1.000 / 1.254	1.000 / 1.247	1.000 / 2.088	1.000 / 2.026
	Free, beta-wavelet init	1.000 / 1.418	1.000 / 1.873	1.000 / 1.082	1.000 / 1.379	1.000 / 1.643	1.000 / 1.319
	Free, random init	2.227 / 2.195	2.054 / 2.439	2.016 / 2.066	2.196 / 2.213	2.020 / 2.871	1.947 / 2.314
	Free, order 32	1.000 / 2.236	1.000 / 2.302	1.000 / 1.394	1.000 / 1.434	1.000 / 2.600	1.000 / 4.165
	Free, order 32, bounded	1.000 / 1.000	1.000 / 1.000	1.000 / 1.000	1.000 / 1.000	1.000 / 1.000	1.000 / 1.000

K complex coefficients per head and layer. We measure the effect of that coordinate system on the optimization in Appendix H.

Shape Versus Scale of the Initial Response. We separate the two in Table 10. Every structured start is rescaled to the same unit supremum on $[0, 2]$, and Table 12 shows that each begins at a supremum of exactly 1.000 with a depth gain of 1.00, so the arms differ only in shape. The two shapes nearest the Schrödinger response, the monotone decay left when its dispersion is removed and the personalized-PageRank response of GPR-GNN, stay within 0.84% of the Schrödinger start on every dataset. The beta-wavelet bands are the outlier among the structured starts. Their composed response begins with a median gain over depth of 1.3×10^{-4} on Minesweeper and Tolokers against 0.39 and 0.16 for the Schrödinger start, so depth annihilates most of their signal there. The arm collapses on three splits in ten on Tolokers and nine in ten on Questions.

Bounded Random Start. A bound alone also does not compensate for a poor shape. Adding the supremum constraint to the noise start leaves it worse than the unbounded noise start on Tolokers and Questions, at 8.90% and 5.90% below the reference against 4.43% and 4.93%, while the constraint helps on Minesweeper, Amazon and inj-Flickr by 0.18% to 0.91% and changes nothing on Elliptic. Its mean over the suite is 2.86% against 2.25% unbounded. The bound of Theorem 1 alone therefore does not explain the advantage of the Schrödinger start, since imposing that bound on a badly shaped start is as likely to hurt as to help.

H Effect of the parameterization on optimization

The results of Appendix G point to conditioning rather than accuracy as the contribution of the Schrödinger form, and we measure that conditioning directly in this appendix. We record four quantities for the constrained response and six free ones over five independent trials and read every one of them twice, once at initialization and once at the epoch validation selects. Both readings appear in Tables 12 and 13, separated by a slash.

H.1 Measured quantities

Parameterization Geometry. The first is the geometry of the map from the parameters the optimizer moves to the responses the network applies, measured in the norm the graph imposes. Writing that

Table 13: Optimization record of each parameterization over five independent trials, where peak curvature is the largest eigenvalue of the training-loss Hessian on the response parameters.

Quantity	Parameterization	Minesweeper	Tolokers	Questions	Amazon	Elliptic	inj-Flickr
Test AUROC (%)	Schrödinger	95.51	85.51	77.08	98.11	99.16	95.12
	Free, Schrödinger init	95.33	84.93	77.25	98.07	99.19	95.27
	Free, monotone decay init	94.65	84.45	77.15	98.11	99.24	95.32
	Free, beta-wavelet init	96.00	80.45	70.18	98.09	99.04	94.29
	Free, random init	94.07	81.20	72.06	97.29	98.72	94.02
	Free, order 32	93.01	83.94	77.30	97.96	99.14	92.83
	Free, order 32, bounded	93.00	84.41	76.87	98.01	99.17	95.20
Initial response gradient	Schrödinger	0.00889	0.00656	0.00206	0.0515	0.0196	5.4e-06
	Free, Schrödinger init	0.437	0.168	0.193	0.713	0.107	0.0153
	Free, monotone decay init	0.436	0.185	0.194	0.76	0.11	0.0153
	Free, beta-wavelet init	0.0103	0.000326	0.13	0.251	0.0433	0.00461
	Free, random init	0.354	0.0853	0.143	0.451	0.0653	0.0122
	Free, order 32	0.763	0.243	0.277	0.929	0.131	0.0223
	Free, order 32, bounded	0.618	0.185	0.227	0.772	0.235	0.0192
Response gradient share	Schrödinger	0.162	0.089	0.041	0.066	0.079	0.126
	Free, Schrödinger init	0.643	0.526	0.521	0.513	0.436	0.889
	Free, monotone decay init	0.624	0.528	0.521	0.503	0.418	0.882
	Free, beta-wavelet init	0.692	0.554	0.447	0.532	0.600	0.788
	Free, random init	0.668	0.531	0.501	0.457	0.407	0.731
	Free, order 32	0.706	0.634	0.637	0.611	0.547	0.770
	Free, order 32, bounded	0.777	0.761	0.735	0.650	0.654	0.871
Peak curvature	Schrödinger	0.00634 / 3.01	0.0128 / 2.75	0.00292 / 0.0265	0.0756 / 0.0448	0.0126 / 0.0208	6.24e-06 / 0.172
	Free, Schrödinger init	2.65 / 15.1	1.84 / 14.4	1.79 / 10.9	4.02 / 3.83	0.607 / 0.53	0.166 / 25
	Free, monotone decay init	2.65 / 17.8	1.9 / 17.6	1.83 / 12.3	4.44 / 4.13	0.625 / 0.635	0.166 / 25.2
	Free, beta-wavelet init	0.0762 / 17.9	0.00339 / 20.2	0.455 / 28.3	1.31 / 3.8	0.261 / 1.01	–
	Free, random init	2.79 / 16.6	0.81 / 9.83	2.36 / 5.74	2.51 / 1.82	0.56 / 0.856	0.131 / 10
	Free, order 32	8.52 / 31.7	3.87 / 24.4	3.72 / 18.7	7.13 / 6.17	1.08 / 0.842	–
	Free, order 32, bounded	–	–	–	–	–	–
Mean curvature	Schrödinger	-0.00016 / 0.227	0.000406 / 0.324	-0.00101 / 0.00546	-0.00074 / 0.0101	-0.000326 / 0.00447	1.98e-06 / 0.0656
	Free, Schrödinger init	0.00552 / 0.103	0.000683 / 0.0838	-0.0158 / 0.305	-0.00789 / 0.0216	0.00322 / 0.00596	-0.000352 / 0.221
	Free, monotone decay init	0.00545 / 0.111	0.000616 / 0.106	-0.0158 / 0.258	-0.00836 / 0.0216	0.00286 / 0.00675	-0.000352 / 0.236
	Free, beta-wavelet init	3.24e-05 / 0.107	7.24e-07 / 0.131	0.00523 / 0.605	-3.57e-05 / 0.0232	0.00184 / 0.00775	–
	Free, random init	0.00202 / 0.0778	-0.000589 / 0.0711	0.00657 / 0.168	-0.00928 / 0.0155	0.00225 / 0.0058	0.000157 / 0.152
	Free, order 32	0.0045 / 0.0787	-0.000543 / 0.0723	-0.0162 / 0.267	-0.00829 / 0.0223	0.00324 / 0.00572	–
	Free, order 32, bounded	–	–	–	–	–	–
Normalized peak curvature	Schrödinger	0.0181 / 2.14	0.0114 / 2.71	-0.00333 / 0.294	0.566 / 0.643	0.0192 / 0.106	-1.44e-05 / 0.701
	Free, Schrödinger init	87.2 / 279	1.3e+03 / 1.91e+04	845 / 8.79e+03	1.54e+03 / 2.84e+03	5.64 / 3.48	1.01e+03 / 8.72e+04
	Free, monotone decay init	88 / 390	1.44e+03 / 7.11e+03	924 / 6e+03	2.15e+03 / 2.86e+03	5.66 / 2.51	1.01e+03 / 9.27e+04
	Free, beta-wavelet init	4.25 / 406	0.75 / 9.35e+03	13.3 / 1.68e+04	624 / 1.75e+03	3.21 / 12.6	0.298 / 1.88e+04
	Free, random init	9 / 428	514 / 7.37e+03	1.41e+03 / 3.61e+03	855 / 636	6.69 / 3.03	223 / 2.43e+04
	Free, order 32	631 / 1.51e+03	1.31e+03 / 8.91e+03	3.76e+03 / 1.74e+04	1.82e+03 / 2.51e+03	16.7 / 14.2	1.02e+03 / 5.49e+04
	Free, order 32, bounded	–	–	–	–	–	–
Normalized mean curvature	Schrödinger	-0.00307 / 0.501	0.000399 / 0.365	-0.0213 / 0.133	-0.261 / 0.12	0.00234 / 0.0245	-3.42e-05 / 0.395
	Free, Schrödinger init	0.174 / 0.924	-1.93 / 66.3	-22.3 / 161	-9.79 / 13	0.00548 / 0.0255	0.376 / 736
	Free, monotone decay init	0.175 / 1.13	-0.566 / 51.7	-34.6 / 115	-9.52 / 12.3	0.00891 / 0.0182	0.6 / 754
	Free, beta-wavelet init	0.00818 / 0.826	0.00126 / 36.8	-5.48 / 252	-1.01 / 5.35	0.00166 / 0.0589	0.00104 / 109
	Free, random init	-0.0157 / 1.15	0.628 / 35.4	1.88 / -26.4	-1.1 / 5.91	0.0129 / 0.0178	-0.57 / 458
	Free, order 32	0.56 / 1.97	-0.498 / 16.9	-95.5 / 220	-6.74 / 8.15	0.00799 / 0.0452	-0.457 / 221
	Free, order 32, bounded	–	–	–	–	–	–

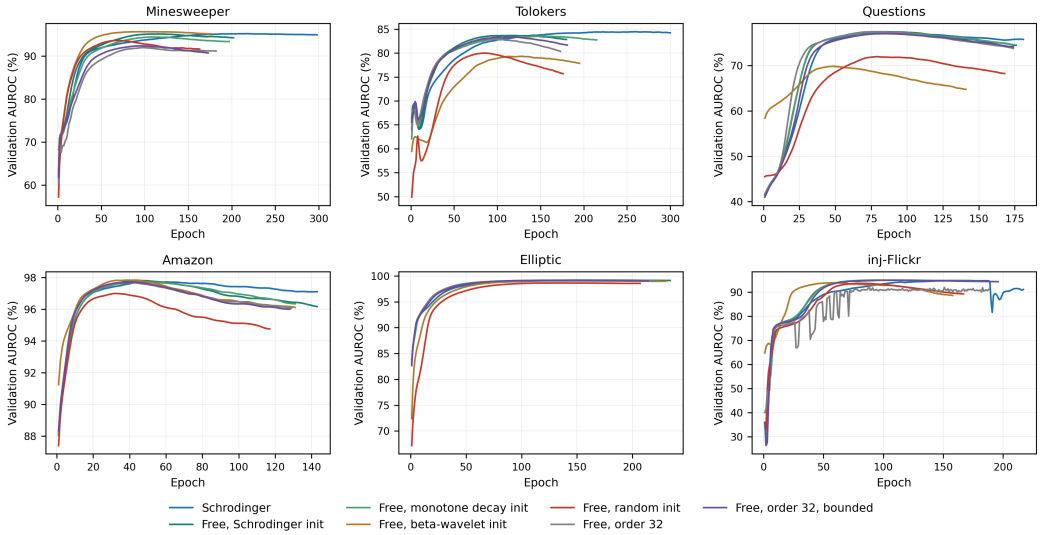

Figure 3: Validation AUROC against epoch for the constrained response and the six free ones, averaged over five independent trials. The order-32 arm on inj-Flickr shows the instability that the collapse count of Table 11 records.

map in the Chebyshev basis turns the inner product $\langle f, g\rangle = N^{-1}\sum_\ell f(\lambda_\ell)g(\lambda_\ell)$ into a fixed Gram matrix that the eigenbasis realization supplies exactly and that Rademacher probing of the sparse operator estimates otherwise. Because each probe reweights the spectrum rather than perturbing the matrix, the estimate is a Gram matrix of the same basis under a measure whose average over the

probes is the spectral one, and it is positive semidefinite by construction. The singular values of the whitened Jacobian then count the coordinate directions the graph can tell apart at a thousandth of the largest. Their ratio over that subspace is the conditioning we report.

Response Profile. The second is the effect of depth on the response. Proposition 2 fixes it for the constrained family at every point of training and leaves it free for every other, so we report the peak and the median of $\prod_t |g_h(\lambda)|$ over a grid on $[0, 2]$ together with the supremum of the single-layer response on the same grid.

Response Gradient. The third is the gradient the optimizer actually receives. We record its norm on the response parameters at the first epoch before clipping and its share of the whole-network gradient norm averaged over training.

Loss Curvature. The fourth is the curvature of the training loss restricted to those parameters, read with the model in evaluation mode so that the loss is deterministic. A Hessian eigenvalue is read in whatever coordinates a parameterization happens to use, and the two families here were developed to use very different ones, so a raw comparison between them measures their scaling as much as their loss surface. We therefore report the curvature twice. The raw reading is in the coordinates the optimizer moves. The normalized reading is the Hessian of the loss taken in response space along the directions each family can reach, scaled so that a unit step changes the response by unit norm under the spectral measure.

Response-Space Normalization. Conjugating the parameter-space Hessian by the whitened Jacobian of the first quantity above would not deliver this, because away from a stationary point the Hessian keeps a term proportional to the gradient times the second derivative of the parameterization, and no conjugation removes it. A free filter is affine in its coefficients, so this term vanishes everywhere and its coefficient Hessian already is the response-space one. For the constrained response we evaluate the same loss through a twin whose Chebyshev coefficients interpolate the constrained response at its current parameters. The twin reproduces the training loss to one part in 10^5 and carries no parameterization term of its own, and we conjugate its Hessian along the constrained tangent directions. The result is curvature per unit change in the function the network applies, invariant under any reparameterization realizing the same responses because the response itself supplies the coordinates. On these graphs the parameter-space conjugation would misstate the constrained reading by up to a factor of 13, which reflects the gradient term. The normalized readings share a Jacobian with the geometry above and are taken on the first split alone.

Dimension-Fair Mean. The largest eigenvalue of the response-space Hessian is a maximum over the identifiable subspace, which is far larger for a free filter than for the constrained response, so we report the trace over that subspace divided by its dimension alongside the peak curvature. That mean is the average curvature over unit-response directions and carries no advantage from dimension.

Hessian Estimation. Peak curvature is the largest eigenvalue of the Hessian block on the response parameters. Power iteration converges to the eigenvalue of largest magnitude, which may have either sign for a nonconvex loss, so a second run shifted by that magnitude reads the largest eigenvalue itself. The mean curvature is a Hutchinson estimate of the block trace divided by the number of parameters in it. Both need Hessian-vector products. We compute them by central differences of the gradient rather than by differentiating it again, because the sparse realization reaches the state through a custom symmetric product and a recomputed Chebyshev recurrence, and neither carries a second derivative. On the eigenbasis realization, where differentiating twice does work, the two agree to one part in a thousand.

Withholding Rule. We withhold some readings rather than report them, and we mark each with a dash in Table 13. A sup-norm projection makes the loss piecewise smooth, and an arm sitting against the boundary is at a point where no second derivative exists. The difference quotient detects that by disagreeing with itself when the displacement is doubled. A mean over the splits that happen to pass that check is a different estimator from a mean over all of them, so one failing split withholds the whole reading instead of shrinking the sample behind it.

H.2 Findings

Identifiability and Conditioning. The geometry of the parameterization favors the constrained response on five of the six graphs. The Schrödinger response is described by 3 to 16 numbers per

dataset, and the graph's own spectral measure identifies 2 to 12 of them. On those five benchmarks the free filter leaves 16 to 256 of its 120 to 960 coefficients unresolved, and its condition number is worse than the constrained one by a factor of 1.3 to 138, at 83 to 373 against 1.2 to 117. On Elliptic the 204K nodes spread the spectral measure evenly enough to resolve all 576 free coefficients, and the free filter reads 4.8 against 13.

Response Profile at Initialization. Every structured start begins at a supremum of exactly 1.000 with a depth gain of 1.00, whether it is the Schrödinger interpolant, the monotone decay, the beta-wavelet bands or the widened filter, so the arms enter training on equal terms. Random coefficients at scale $K^{-1/2}$ are the only start that violates both, at a supremum of 1.95 to 2.23 and a depth gain that reaches 6.26 on Tolokers, and the failure of that arm in Appendix G is therefore visible before any training happens.

Response Profile After Training. The response profile separates the families at the end of training. At the selected epoch the Schrödinger response still reads exactly 1.000 and 1.00, as Theorem 1 and Proposition 2 guarantee at every point of training. Every free response has left the unit ball by then, at a supremum of 1.08 to 4.17, and the modulus its T layers compose reaches 8.1 on Tolokers at matched degree and 52.3 from the beta-wavelet start. The one arm the projection holds at 1.000 moves in the opposite direction instead and composes to 0.17 on Minesweeper, so depth annihilates its signal. Free filters therefore realize both failure modes of Appendix G whereas Proposition 2 rules out both for the constrained response. Figure 3 shows the validation trajectories these profiles produce.

Raw Peak Curvature. The curvature separates the families on every dataset. In raw coordinates the largest eigenvalue at initialization is 6.2×10^{-6} to 0.076 for the constrained family against 0.13 to 8.52 for every free one except the beta-wavelet start, which is flat for a degenerate reason at 0.0034 to 1.31 since its composed response has already annihilated the signal. Training narrows that gap without closing it, from 0.021 to 3.01 against 0.53 to 32, the smallest per-dataset ratio being 3.6 on Tolokers and the largest 216 on Questions. None of those numbers is invariant to a change of coordinates.

Normalized Peak Curvature. The normalized reading is invariant and separates the families more widely. The constrained response has the smallest curvature per unit change in the response of any arm on all six datasets and at both ends. Against the nearest free arm the factor runs from 66 to 2.1×10^4 at initialization and from 24 to 2.7×10^4 at the selected epoch. This advantage does not follow from the conditioning. Elliptic is the one graph whose spectral measure resolves the free filter better than the constrained one, and the constrained response is still the flattest arm there by a factor of 167 at initialization and 24 at convergence.

Normalized Mean Curvature. The dimension-fair mean curvature favors the constrained response on five of the six datasets. At the selected epoch it puts that response below every free arm there, by factors of 1.6 on Minesweeper to 276 on inj-Flickr, and on Elliptic it sits at 0.025 against 0.018 to 0.059. At initialization the mean again favors the constrained response on five of the six datasets, while several free arms are strongly negative, meaning a random unit-response direction is concave rather than flat. The constrained parameterization therefore has a flatter worst direction per unit change in the response on every graph and a flatter average direction on five of the six at either end.

Withheld Readings. We withhold three readings. The sup-norm projection places the bounded wide filter on its own constraint boundary, where the response already attains the bound and no second derivative exists, and the difference quotient disagrees with itself under doubling on splits of all six datasets, so this arm carries no curvature anywhere. The beta-wavelet start on inj-Flickr fails the same check at the selected epoch, as does the order-32 filter on the same graph, where one split in five collapses to 83.83% AUROC against roughly 95% on the other four and invalidates the difference quotient. That last case is the instability of Appendix G observed a second time. The split it affects here is not the one it affected there, which indicates that the failure is stochastic rather than tied to particular splits.

I SENSITIVITY ANALYSIS

We vary one hyperparameter of the propagation bank at a time around the validated configuration of each ablation dataset in Table 14 while every other setting stays frozen. The three axes are the number

Table 14: Hyperparameter sensitivity in terms of test AUROC over ten independent trials. Bold marks the validated configuration of each dataset, which is held fixed on every axis but the one that varies.

Setting	Minesweeper	Tolokers	Questions	Amazon	Elliptic	inj-Flickr
Spectral heads H						
1	94.93±0.28**	84.71±0.66**	**76.87±1.21**	98.05±0.30	99.17±0.13	**95.10±0.15**
2	94.99±0.39**	84.83±0.67*	76.63±1.07	98.04±0.33	99.25±0.15	95.19±0.18
4	**95.55±0.56**	**85.08±0.76**	76.73±1.00	**98.01±0.32**	**99.24±0.12**	95.20±0.17*
8	95.44±0.66	85.02±0.72	76.77±1.07	98.03±0.32	99.24±0.10	95.17±0.21
Depth T at fixed total time						
1	91.76±0.43**	83.66±0.88**	76.42±1.19*	97.89±0.31*	98.95±0.10**	93.97±0.20**
2	93.76±0.27**	84.59±0.58**	76.79±1.05	97.93±0.34	99.18±0.08	94.72±0.24**
4	95.40±0.39	84.86±0.60	**76.87±1.21**	**98.01±0.32**	**99.24±0.12**	94.83±0.30**
8	**95.55±0.56**	**85.08±0.76**	76.07±1.29**	98.02±0.39	99.27±0.08	**95.10±0.15**
16	95.95±0.85	84.83±0.68	75.87±1.14**	98.02±0.39	99.28±0.10	93.00±6.03
Total-time scale						
×0.25	94.19±0.43**	84.53±0.62**	75.20±0.92**	98.04±0.35	99.19±0.08	93.17±4.12**
×0.5	94.67±0.35**	84.67±0.78**	76.45±0.95	98.00±0.36	99.21±0.10	94.92±0.30*
×1	**95.55±0.56**	**85.08±0.76**	**76.87±1.21**	**98.01±0.32**	**99.24±0.12**	**95.10±0.15**
×2	95.84±0.68	85.11±0.71	76.62±0.90	97.93±0.31	99.20±0.09	94.79±0.23**
×4	96.60±0.58**	84.91±0.85	76.00±1.07*	97.82±0.38	99.05±0.09**	94.71±0.20**

of spectral heads, the depth at fixed total evolution time and the total evolution time at the validated depth. We run every axis on the ten evaluation splits under the paired protocol of Section 4.3, so asterisks carry the same meaning as in the ablation tables.

Spectral Heads. Minesweeper and Tolokers are the two graphs that need more than one head, and collapsing the bank to a single head costs 0.62% and 0.37%. Both saturate at four heads, and eight heads never produce a resolved change anywhere on the suite. The other four graphs are flat along this axis to within 0.23%. On Questions no multi-head setting outperforms the validated single head, so splitting the channels neither helps nor hurts a graph whose dynamics are one strongly damped regime.

Depth at Fixed Total Time. A single step removes all interleaving between propagation and modulation, and it costs from 0.12% on Amazon to 3.79% on Minesweeper with a resolved loss on all six graphs. Sixteen steps shorten each linear flow and multiply the modulation and mixing applications. That setting costs a resolved 0.99% only on Questions and widens the spread over splits on inj-Flickr to 6.03%. A plateau sits between the two ends, from four to sixteen steps on Minesweeper and Tolokers, two to sixteen on Amazon and Elliptic and two to four on Questions, while inj-Flickr peaks at its validated eight steps.

Total Evolution Time. Quartering the total evolution time is resolved on four of the six graphs and costs from 0.55% on Tolokers to 1.93% on inj-Flickr, so too little transport is the common failure. Lengthening it has opposite effects across graphs. Minesweeper gains 1.06% at four times the validated budget, which is consistent with a sparse high-frequency signal that requires more transport than the searched grid allowed, whereas Questions, inj-Flickr and Elliptic lose at most 0.87% at the same setting.

Validated Versus Test-Optimal Settings. The validated configuration is not the test-optimal setting on every axis. Minesweeper sits below the test optimum in both depth and total time, and on inj-Flickr four heads gain a resolved 0.10% over the validated single head. Both gaps follow from selecting on validation data alone, which is the rule every method in the comparison follows. We report the validated configurations throughout and do not re-search the three axes, so where these sweeps disagree with the validated setting, the main tables understate what the architecture can reach.

J Search protocol and validated configurations

Baseline Selection. The comparison set is the five published spectral detectors of Section 2 plus one representative of each generic family they extend: a fixed low-pass operator in GCN, a preserved

Table 15: Validated per-dataset configurations of G-SNLS, chosen on validation data by the two-stage search of this appendix and then frozen across all ten evaluation splits.

Dataset	d	T	Δt	H	Context	Dispersion	Readout	Potential	LR	WD	Dropout	η
Tolokers	64	8	0.75	4	raw	quadratic	intensity	linear	3e−3	1e−4	0.2	0
Questions	128	4	0.75	1	multi-scale	quadratic	signed	MLP	1e−3	1e−3	0.5	1
Amazon	128	4	0.75	4	homophily	quadratic	signed	MLP	3e−3	1e−4	0.2	1
T-Finance	128	8	0.5	4	raw	linear	signed	linear	3e−3	0	0.2	1
Elliptic	128	4	0.75	4	raw	quadratic	signed	linear	3e−3	1e−4	0.5	1
inj-Amazon	256	8	0.2	4	neighbor	quadratic	intensity	linear	1e−3	5e−4	0.5	1
Minesweeper	64	8	0.2	4	homophily	quadratic	signed	MLP	1e−2	5e−4	0.2	1
inj-Flickr	128	8	0.75	1	raw	linear	signed	linear	1e−2	0	0.2	0

Table 16: Measured seconds per training epoch and peak allocated GPU memory in gibibytes, on one device and one library stack under the validated configuration of each method.

	GCN		GraphSAGE		GPR-GNN		AMNet		G-SNLS	
Dataset	s/epoch	GiB	s/epoch	GiB	s/epoch	GiB	s/epoch	GiB	s/epoch	GiB
Tolokers	0.019	1.10	0.015	1.11	0.015	0.14	0.029	0.58	0.063	1.18
Questions	0.016	0.97	0.006	0.59	0.011	0.19	0.019	0.42	0.169	3.05
Amazon	0.278	17.33	0.125	8.64	0.088	0.87	1.077	17.28	0.050	1.46
T-Finance	0.767	42.83	0.339	21.16	0.441	4.07	5.447	83.00	2.872	4.63
Elliptic	0.042	2.29	0.035	2.19	0.017	0.92	0.356	5.03	1.184	11.47
inj-Amazon	0.019	1.15	0.016	1.62	0.011	0.16	0.069	1.27	0.144	2.52
Minesweeper	0.005	0.23	0.002	0.17	0.009	0.04	0.013	0.33	0.040	0.85
inj-Flickr	0.041	2.57	0.038	2.29	0.017	0.66	0.241	3.38	0.748	8.50

self term in GraphSAGE, a learned polynomial response in GPR-GNN and a continuous dynamical system in GRAND. The oscillator and reaction-diffusion architectures cited alongside GRAND are node classifiers rather than detectors and are represented by GRAND, the diffusion baseline that the anomaly-detection literature itself uses.

Search Protocol. Our two-stage search screens candidate configurations on the first development split and then rescreens the best three on both development splits using validation data only. G-SNLS screens forty-eight sampled configurations per dataset because its space includes discrete architectural choices. A baseline screens its complete grid when that grid is smaller, which is four configurations for GHRN because it inherits its backbone from BWGNN, and screens eighteen or nineteen sampled configurations otherwise with its published default added when the sample misses it. The slowest runs on the two densest graphs drop to between eight and thirteen. The macro F1 decision threshold comes from the grid $\{0.05, 0.10, \ldots, 0.95\}$ on validation data and is carried unchanged to the test set for every method.

Validated Configurations. We list the validated configuration of G-SNLS for every dataset in Table 15. The columns give hidden width d, depth T, base step Δt, head count H, input context, dispersion order, readout, potential network, learning rate, weight decay, dropout and focal exponent η. Every dataset uses a single-layer encoder and decoder with one shared complex channel map and the same step at every layer. Initial values of the per-head parameters follow the validated recipes stored with the code, and every method trains with label smoothing 0.05 and an exponential moving average with decay 0.9 for evaluation.

K Training cost

Timing Protocol. We report the seconds per training epoch and the peak GPU memory of every method in Table 16 and list the parameter counts of the same runs in Table 17. The timing covers the forward and backward passes together with the evaluation pass that the shared protocol runs at every epoch. We time each method under its validated configuration and include only the four baselines that share the library stack of G-SNLS because timings across different stacks are not comparable.

Sparse Graphs. G-SNLS runs deeper than GCN and GraphSAGE and carries two real tensors per complex channel, so it spends the most time and memory per epoch on the six sparse graphs. The

Table 17: Trainable parameters in thousands under the validated configuration of each method and dataset, counted on the same runs that Table 16 times.

Dataset	GCN	GraphSAGE	GPR-GNN	AMNet	G-SNLS
Tolokers	26	203	3	13	10
Questions	176	51	19	32	513
Amazon	106	210	7	205	76
T-Finance	26	52	3	201	37
Elliptic	141	282	43	241	76
inj-Amazon	296	590	99	395	918
Minesweeper	101	201	1	200	17
inj-Flickr	227	454	129	327	162

model evaluates its propagator once per layer for all heads together, so the head count does not multiply its cost.

Dense Graphs. The propagator is a function of the Laplacian rather than a gather over edges, so its working set scales with Nd instead of $|\mathcal{E}|d$. On the two densest graphs G-SNLS is therefore no longer the slowest method. It is the fastest of the five on Amazon, where a cached eigendecomposition turns the propagator into two dense products, and it runs faster than AMNet on T-Finance. G-SNLS also needs far less peak memory than GCN, GraphSAGE and AMNet on both graphs, since each of these baselines materializes a tensor of shape $|\mathcal{E}| \times d$ inside every propagation step. GPR-GNN also applies its filter as a sparse matrix product and stays below G-SNLS in memory on every graph.

Model Size. Parameter counts follow the validated widths and track neither ordering, so model size explains neither the speed nor the memory results.