Oversmoothing as Representation Degeneracy in Neural Sheaf Diffusion

Arif Dönmez[1,2,*] **Axel Mosig**[2,3] **Ellen Fritsche**[1,2,4] **Katharina Koch**[1,2]
[1] IUF – Leibniz Research Institute for Environmental Medicine, Düsseldorf, Germany
[2] DNTOX GmbH, Düsseldorf, Germany
[3] Bioinformatics Group, Ruhr University Bochum, Bochum, Germany
[4] Swiss Centre for Applied Human Toxicology (SCAHT), Basel, Switzerland
[*] Corresponding author: arif.doenmez@ruhr-uni-bochum.de

Abstract

Neural Sheaf Diffusion (NSD) learns a task-adapted sheaf geometry for node classification. We connect the study of these learned geometries with Geometric Invariant Theory (GIT) to analyse their representation-theoretic structure and develop mechanisms for geometric control during training. Identifying cellular sheaves with representations of the incidence quiver connects direct-sum decompositions to the harmonic space and motivates the study of trivial or low-complexity summands as potential degeneration modes. We make this perspective concrete by focusing on the trivial representation: on an embedded copy, the sheaf Dirichlet energy reduces to weighted scalar graph energy. The GIT stability framework provides a criterion for excluding such copies within the semistable class. Using its moment-map characterisation of polystability, we construct differentiable surrogates that encourage the corresponding balance conditions. We identify an equal-stalk obstruction to assigning a destabilising weight to the trivial representation and show how nonuniform stalk dimensions remove it. Controlled energy-minimisation experiments reproduce trivial-component formation, while native feature-conditioned NSD experiments on heterophilic graphs show how moment-map regularisation changes the balance and Dirichlet-energy trajectories of learned geometries during training, alongside their classification behaviour.

1 Introduction

Many Graph Neural Networks (GNNs) can be understood as discrete diffusion processes on graphs [Kipf and Welling, 2017, Veličković et al., 2018]. In message-passing architectures, nodes repeatedly aggregate information from their neighbors, effectively implementing a form of Laplacian smoothing. While this diffusion perspective explains the success of GNNs in propagating local information, it also exposes a fundamental limitation: repeated smoothing can drive node representations toward indistinguishable limits. This phenomenon, known as oversmoothing, is especially problematic in deep architectures where node features collapse to low-dimensional or nearly constant subspaces [Li et al., 2018, Chen et al., 2020]. In the classical unnormalized graph Laplacian setting, this collapse is a structural consequence of the underlying diffusion geometry: the heat equation asymptotically projects features onto the kernel of the Laplacian, which for a connected graph consists of constant signals. However, this tendency toward identical node representations can be mismatched for heterophilic graphs, motivating the search for richer geometric structures that avoid collapsing into overly simple harmonic states.

Neural Sheaf Diffusion (NSD) addresses this by lifting the diffusion process into a vector bundle-like structure, replacing the standard Laplacian with a sheaf Laplacian whose restriction maps are learned from the data [Hansen and Ghrist, 2019, Bodnar et al., 2022]. By learning a task-adapted

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

geometry, NSD allows the network to model heterophily: the space of global sections, which governs the limit of fixed linear sheaf diffusion, can encode complex, signed, and multi-dimensional feature agreements rather than simple constants. However, optimizing these highly parameterized general sheaves can be challenging [Barbero et al., 2022b,a]. Furthermore, gradient-based learning exhibits implicit biases [Rahaman et al., 2019, Shah et al., 2020, Arora et al., 2019]. This motivates investigating whether unregularized learning can favor overly simple restriction maps and thereby limit the expressive geometry available to diffusion.

In this work, we study this potential failure mode through the lens of quiver representation theory [Assem et al., 2006, Schiffler, 2014] and its geometric interpretation [Bongartz, 1998, Kraft, 1982, Derksen and Weyman, 2005, Reineke, 2008]. Identifying cellular sheaves on graphs with representations of the associated incidence quiver realizes the learned restriction maps as points in a finite-dimensional representation space. This provides a representation-theoretic perspective on asymptotic sheaf diffusion: direct-sum decompositions of the sheaf representation induce corresponding decompositions of its harmonic space. We investigate trivial or low-complexity summands as potential degeneration modes, focusing on the trivial representation, $\mathcal{F}_{\mathrm{triv}}$, on whose embedded copies the sheaf Dirichlet energy reduces to weighted scalar graph energy.

To control such potential degeneration during learning, we invoke Geometric Invariant Theory (GIT) [Kraft, 1985, Popov and Vinberg, 1994, Mumford et al., 1994], particularly King's stability framework for quiver representations [King, 1994]. Its stability framework provides a criterion for excluding the trivial representation as a subrepresentation when its stability weight is negative. The moment-map characterization of polystability motivates differentiable regularizers that encourage the corresponding balance conditions in the learned restriction maps. This yields a conditional exclusion mechanism: if the learned geometry satisfies the corresponding polystability conditions, a negative weight rules out the trivial subrepresentation. Crucially, our algebraic analysis identifies a structural obstruction in equal-stalk architectures: when vertex and edge stalks have equal dimensions ($d_v = d_e$), admissibility forces the weight of the trivial all-object representation to vanish, placing it on a stability wall. Nonuniform stalk dimensions remove this obstruction, allowing adaptive stability parameters to assign it a destabilizing weight.

Our experiments examine this mechanism through controlled energy minimization, recovery from collapsed initial configurations, and native feature-conditioned NSD training. Controlled Dirichlet-energy minimization demonstrates the formation of trivial direct summands, while architectural comparisons confirm that rectangular stalks allow learned stability weights to leave the wall imposed by equal stalk dimensions. In native NSD training, moment-map regularization steers the balance and Dirichlet-energy trajectories of the learned geometries, with effects depending on the regularizer, dataset and network depth. Recovery experiments investigate learning from explicitly constructed trivial components with initially collapsed node signals. Under increased classification difficulty and on the WebKB graphs studied, these experiments provide evidence of improved held-out accuracy under regularization, including at fixed training endpoints where both regularized and unregularized models achieve complete training accuracy. Together, these findings illustrate geometric control during learning and setting-dependent generalization benefits, without establishing that the task improvements result from removal of trivial components or constitute a general prevention of oversmoothing.

Contributions. Our main contribution is to connect the study of Neural Sheaf Diffusion with Geometric Invariant Theory. Through a quiver-theoretic interpretation of learned sheaf geometries, stability, polystability and moment maps provide both a finer analysis of potential degeneration modes and concrete mechanisms for geometric control during learning. We develop differentiable regularizers and identify architectural conditions that enable this control, with experiments demonstrating their influence on learned geometries.

2 Background

Graph Diffusion and Oversmoothing. Let $G = (V, E)$ be an undirected graph with $n = |V|$ nodes, and let $\mathbf{X} \in \mathbb{R}^{n \times d}$ denote a node signal matrix. Standard graph diffusion is governed by the heat equation $\dot{\mathbf{X}}(t) = -\Delta\mathbf{X}(t)$, where Δ is the normalized graph Laplacian. As $t \to \infty$, the signal converges to the orthogonal projection onto $\ker(\Delta)$. For a connected graph, the scalar graph Laplacian has a one-dimensional kernel corresponding to globally constant signals, up to the usual

degree weighting in the symmetrically normalized case. Thus, fixed linear diffusion suppresses all signal components orthogonal to this kernel, providing the basic smoothing mechanism underlying oversmoothing analyses in GNNs [Li et al., 2018, Chen et al., 2020].

Cellular Sheaves. A *cellular sheaf* $\mathcal{F}$ over a graph G equips the network with a richer structure by assigning a vector space $\mathcal{F}(v)$ to each vertex $v \in V$, a vector space $\mathcal{F}(e)$ to each edge $e \in E$, and a linear restriction map $\mathcal{F}_{v \trianglelefteq e} : \mathcal{F}(v) \to \mathcal{F}(e)$ for every incidence $v \trianglelefteq e$ [Curry, 2014, Shepard, 1985]. The space of 0-cochains, or vertex-level sheaf signals, is $C^0(G;\mathcal{F}) = \bigoplus_{v \in V} \mathcal{F}(v)$, and the space of 1-cochains is $C^1(G;\mathcal{F}) = \bigoplus_{e \in E} \mathcal{F}(e)$. Choosing an arbitrary edge orientation $e = (u, v)$, the coboundary operator $\delta_{\mathcal{F}} : C^0(G;\mathcal{F}) \to C^1(G;\mathcal{F})$ is defined component-wise as $(\delta_{\mathcal{F}}\mathbf{x})_e = \mathcal{F}_{v \trianglelefteq e}\mathbf{x}_v - \mathcal{F}_{u \trianglelefteq e}\mathbf{x}_u$. It measures the disagreement of neighboring vertex data after both have been transported into the corresponding edge stalk. Equipping the stalks with inner products induces inner products on the cochain spaces.

Sheaf Laplacians and NSD. The sheaf Laplacian is defined as $\Delta_{\mathcal{F}} = \delta_{\mathcal{F}}^* \delta_{\mathcal{F}}$ [Hansen and Ghrist, 2019]. It is positive semidefinite, and its kernel coincides exactly with the space of *global sections* $H^0(G;\mathcal{F}) = \ker(\delta_{\mathcal{F}}) = \ker(\Delta_{\mathcal{F}})$. A global section represents an assignment of local vertex data that is perfectly consistent across all edges after transport through the sheaf maps. Neural Sheaf Diffusion (NSD) generalizes classical graph diffusion by learning feature-conditioned restriction maps through a parameterized sheaf learner [Bodnar et al., 2022]. For a fixed sheaf geometry, the linear diffusion equation $\dot{\mathbf{X}}(t) = -\Delta_{\mathcal{F}}\mathbf{X}(t)$ converges asymptotically to $\Pi_{H^0(G;\mathcal{F})}\mathbf{X}(0)$. This identifies global sections as the limiting signal space of fixed linear sheaf diffusion and motivates studying their algebraic structure. For an invertible positive block normalization N, as used in our implementation, the normalized operator $N\Delta_{\mathcal{F}}N$ instead has kernel $N^{-1}H^0(G;\mathcal{F})$; the harmonic coordinates are transported accordingly (Appendix A).

Quiver Representations. To analyze this algebraic structure, we use quiver representation theory [Derksen and Weyman, 2005, Reineke, 2008, Schiffler, 2014]. A quiver $Q = (Q_0, Q_1)$ consists of vertices Q_0 and directed arrows Q_1. A representation M of Q assigns a vector space M_i to each $i \in Q_0$ and a linear map $M_a : M_{s(a)} \to M_{t(a)}$ to each arrow $a \in Q_1$, where $s(a)$ and $t(a)$ denote source and target. For a fixed dimension vector $\mathbf{d} = (\dim M_i)_{i \in Q_0}$, the space of all such representations is $\mathrm{Rep}(Q, \mathbf{d}) = \prod_{a \in Q_1} \mathrm{Hom}\left(k^{d_{s(a)}}, k^{d_{t(a)}}\right)$. The base-change, or gauge, group $G_{\mathbf{d}} = \prod_{i \in Q_0} GL(d_i)$ acts on this space by changing bases in the assigned vector spaces. We use the identification of cellular sheaves on a graph with representations of its incidence quiver to connect learned sheaf geometries with direct-sum decompositions, stability and moduli of quiver representations.

3 Cellular Sheaves as Quiver Representations

To analyze the representation geometry of Neural Sheaf Diffusion, we first make explicit the dictionary between cellular sheaves on a graph and quiver representations. This correspondence identifies the learnable restriction maps of an NSD model with representation data in a finite-dimensional algebraic space.

The Incidence Quiver. For an undirected graph $G = (V, E)$, define its *incidence quiver* $Q_G = (Q_0, Q_1)$ as the bipartite quiver with vertex set $Q_0 = V \sqcup E$. For every incidence $v \trianglelefteq e$, we introduce one directed arrow $a_{v,e} : v \to e$. Thus, $Q_1 = \{a_{v,e} : v \to e \mid v \trianglelefteq e\}$. The quiver Q_G should not be confused with an orientation of the original graph: graph edges become objects of the quiver, and each such edge-object receives arrows from its incident endpoints. For a graph edge $e = \{u, v\}$, the incidence quiver therefore contains two arrows $u \to e$ and $v \to e$. All arrows run from vertex-objects to edge-objects; consequently, Q_G is acyclic, even when the underlying graph contains cycles.

$$\underbrace{u \text{ — } v}_{G\text{: } e=\{u,v\}} \quad \longmapsto \quad \underbrace{u \longrightarrow e \longleftarrow v}_{Q_G} \quad \longmapsto \quad \underbrace{\mathcal{F}(u) \xrightarrow{A_{u,e}} \mathcal{F}(e) \xleftarrow{A_{v,e}} \mathcal{F}(v)}_{\text{stalks and restriction maps}}.$$

The Correspondence. A representation M of Q_G assigns a vector space M_v to each graph vertex $v \in V$, a vector space M_e to each graph edge $e \in E$, and a linear map $M_{a_{v,e}} : M_v \to M_e$ to each incidence arrow. This is exactly the data of a cellular sheaf $\mathcal{F}$ on G: we set $M_v = \mathcal{F}(v)$,

$M_e = \mathcal{F}(e)$, and $M_{a_{v,e}} = \mathcal{F}_{v \trianglelefteq e}$. Conversely, every representation of Q_G defines a cellular sheaf by the same assignment. Hence the category of finite-dimensional cellular sheaves on G is equivalent to the category of finite-dimensional representations of the incidence quiver Q_G.

Dimension Vectors and Parameter Space. In Neural Sheaf Diffusion, the stalk dimensions are architectural choices. Writing $d_v = \dim \mathcal{F}(v)$ and $d_e = \dim \mathcal{F}(e)$ gives a dimension vector $\mathbf{d} = \big((d_v)_{v \in V}, (d_e)_{e \in E}\big)$. For this fixed dimension vector, the space of cellular sheaves with these stalk dimensions is the quiver representation space $\mathrm{Rep}(Q_G, \mathbf{d}) = \prod_{v \trianglelefteq e} \mathrm{Hom}\left(k^{d_v}, k^{d_e}\right)$. For each layer and forward pass, the feature-conditioned sheaf learner produces a point in $\mathrm{Rep}(Q_G, \mathbf{d})$. Training updates the learner parameters and thereby moves the generated restriction maps through the subset of representation space accessible to this parameterization.

Gauge Equivalence and Moduli. The matrix representation of a sheaf depends on choices of bases in the stalks. Changing bases by $g_v \in GL(d_v)$ and $g_e \in GL(d_e)$ transforms each restriction map by $\mathcal{F}_{v \trianglelefteq e} \longmapsto g_e \mathcal{F}_{v \trianglelefteq e} g_v^{-1}$. This is precisely the base-change action of the gauge group $G_{\mathbf{d}} = \prod_{v \in V} GL(d_v) \times \prod_{e \in E} GL(d_e)$ on $\mathrm{Rep}(Q_G, \mathbf{d})$. Thus, gauge orbits classify the underlying sheaf representations up to isomorphism. The diffusion operator additionally depends on the inner products on the stalks, which must be transported consistently under changes of bases. This leads naturally to a moduli problem for the underlying representations, for which GIT provides stability-dependent quotient constructions.

4 The Kernel Decomposition Theorem

Having established the equivalence between cellular sheaves and representations of the incidence quiver Q_G, we can recast the diffusion limit in the language of homological algebra.

Global Sections as Categorical Limits. For a cellular sheaf $\mathcal{F}$ viewed as a representation diagram $M : Q_G \to \mathbf{Vect}$, the space of global sections is canonically isomorphic to the categorical limit of this diagram. Indeed, the limit consists of compatible assignments to all vertex and edge stalks, while $H^0(G; \mathcal{F}) = \ker(\delta_{\mathcal{F}})$ records the corresponding vertex assignments satisfying the compatibility equations. Thus the operation sending a sheaf to its harmonic space is the global section functor $\Gamma : \mathrm{Rep}(Q_G) \to \mathbf{Vect}$, $\mathcal{F} \longmapsto H^0(G; \mathcal{F})$. Equivalently, Γ is the limit functor applied to the incidence-quiver representation. As a limit functor, it is left-exact; in particular, it preserves finite limits.

Functorial Decomposition. By the Krull–Schmidt theorem, every finite-dimensional representation in the abelian category $\mathrm{Rep}(Q_G)$ admits a direct-sum decomposition into indecomposable objects, unique up to isomorphism and permutation: $\mathcal{F} \cong \bigoplus_{k=1}^{K} \mathcal{F}^{(k)}$ [Assem et al., 2006]. Since finite direct sums in $\mathbf{Vect}$ are biproducts, and since Γ preserves finite products, the global section functor preserves this decomposition up to natural isomorphism. Applying Γ gives $H^0(G; \mathcal{F}) \cong \bigoplus_{k=1}^{K} H^0\big(G; \mathcal{F}^{(k)}\big)$. Thus the decomposition of the diffusion limit does not require a separate spectral argument or manual block-diagonalization of the sheaf Laplacian. It is a functorial consequence of viewing learned cellular sheaves as objects of the incidence-quiver representation category.

Homological Interpretation of Oversmoothing. For a fixed sheaf geometry, continuous linear diffusion asymptotically projects node features onto the global-section space $H^0(G; \mathcal{F})$. Since the global section functor preserves finite direct sums, this limiting signal space decomposes algebraically into the harmonic contributions of the individual Krull–Schmidt summands. This connects asymptotic diffusion to representation structure and motivates investigating trivial or low-complexity summands as potential degeneration modes. Their effect on task-relevant information depends on their embeddings in the vertex stalks and on how the input features project onto the resulting harmonic space. We use this homological perspective to motivate geometric control of learned representations through stability and moment-map regularization.

Subrepresentations and Harmonic Injection. While the Krull–Schmidt theorem describes direct-sum decompositions, the relationship between subrepresentations and global sections is governed by the exactness properties of the limit functor.

Lemma 4.1 (Harmonic Injection)**.** *Let $\mathcal{F}$ be a cellular sheaf on G and let $\mathcal{F}' \hookrightarrow \mathcal{F}$ be a subrepresentation of the corresponding incidence quiver. Then there is an induced injection of global sections $H^0(G; \mathcal{F}') \hookrightarrow H^0(G; \mathcal{F})$.*

Proof. The global section functor $\Gamma : \mathrm{Rep}(Q_G) \to \mathbf{Vect}$ acts by taking the categorical limit of the representation diagram. Because the limit functor is left exact, it preserves monomorphisms. Therefore, the inclusion of the subrepresentation $\mathcal{F}' \hookrightarrow \mathcal{F}$ induces an injective linear map $\Gamma(\mathcal{F}') \hookrightarrow \Gamma(\mathcal{F})$. □

This extends the decomposition perspective beyond direct summands: subrepresentations contribute injectively to the global-section space even when they do not split off. The next section uses this observation to motivate stability-based control of potential degeneration modes.

5 Representation Stability and Oversmoothing

Having developed a representation-theoretic perspective on potential oversmoothing mechanisms, we now turn to stability and moduli as a framework for studying and controlling learned sheaf geometries.

The Trivial Subrepresentation. To understand why representation degeneracy relates to oversmoothing, consider the *trivial representation* $\mathcal{F}_{\mathrm{triv}}$, which assigns a one-dimensional space to every vertex and edge and the identity map to every incidence. If $\mathcal{F}_{\mathrm{triv}}$ exists as a subrepresentation of a learned sheaf $\mathcal{F}$, there exists a 1-dimensional subspace $W_v \subset \mathbb{R}^{d_v}$ at every vertex v and a 1-dimensional subspace $W_e \subset \mathbb{R}^{d_e}$ at every edge e such that the restriction maps act as an isomorphism between these subspaces. Since this subrepresentation is isomorphic to $\mathcal{F}_{\mathrm{triv}}$, we may choose compatible bases $w_v \in W_v$ and $w_e \in W_e$ so that $\mathcal{F}_{v \trianglelefteq e} w_v = w_e$. For any signal $\mathbf{x}$ restricted to this subrepresentation, the node features take the form $\mathbf{x}_v = c_v w_v$ for some scalars $c_v \in \mathbb{R}$. The sheaf Dirichlet energy along an edge $e = (u, v)$ then evaluates to:

$$\begin{aligned} \|\mathcal{F}_{v \trianglelefteq e}\mathbf{x}_v - \mathcal{F}_{u \trianglelefteq e}\mathbf{x}_u\|^2 &= \|c_v(\mathcal{F}_{v \trianglelefteq e} w_v) - c_u(\mathcal{F}_{u \trianglelefteq e} w_u)\|^2 \\ &= \|c_v w_e - c_u w_e\|^2 \\ &= (c_v - c_u)^2 \|w_e\|^2. \end{aligned}$$

Thus, on the embedded trivial copy, the sheaf Dirichlet energy reduces to weighted scalar graph energy. For a connected graph, its global sections have constant coefficients c_v in the compatible bases.

Optimization Bias and Collapse. Implicit biases toward simple solutions provide a motivation for studying low-complexity representation structures in learned sheaf geometries [Rahaman et al., 2019, Gunasekar et al., 2017, Shah et al., 2020]. The trivial representation offers a concrete candidate: signals with constant coefficients in its compatible bases have zero Dirichlet energy. If learning favors such components and the node signals concentrate in them without preserving task-relevant distinctions, scalar-like smoothing can dominate the resulting representation. We investigate this as a potential degeneration mechanism rather than a universal consequence of classification training.

Geometric Intuition for Degeneration. Quiver representation theory provides a geometric language for degeneration through nonclosed gauge orbits and their boundary representations [Bongartz, 1998, Reineke, 2008]. A useful mental model in our setting is a learned sheaf whose decomposition becomes dominated by copies of $\mathcal{F}_{\mathrm{triv}}$, with $\bigoplus \mathcal{F}_{\mathrm{triv}}$ as an idealized limiting configuration. On each trivial copy, the sheaf Dirichlet energy reduces to weighted scalar graph energy, and harmonicity requires constant coefficients across a connected graph. If the learned signals concentrate in these components without retaining task-relevant distinctions, the richer sheaf geometry effectively supports only scalar-like smoothing. This motivates studying trivial-component formation and mechanisms for controlling the learned representation geometry, without assuming that every trivial component is harmful or that classification training necessarily converges to this configuration. Our experiments in Section 7 examine this perspective through controlled trivial-component formation, recovery from deliberately collapsed initial configurations, and geometric regularization during native NSD training. These complementary settings illustrate how the proposed representation-theoretic viewpoint informs both the analysis and control of learned sheaf geometries.

Representation Degeneracy. By the kernel decomposition theorem, any direct-sum decomposition $\mathcal{F} \cong \bigoplus_k \mathcal{F}^{(k)}$ induces a decomposition $H^0(G;\mathcal{F}) \cong \bigoplus_k H^0(G;\mathcal{F}^{(k)})$. This makes the harmonic contributions of individual summands explicit. We use *representation degeneracy* to describe configurations in which the surviving signal is dominated by contributions that fail to retain task-relevant

distinctions. Trivial or low-complexity summands provide concrete candidates for this mechanism, although their presence alone does not imply an uninformative diffusion limit.

King Stability. King stability provides an algebraic way to exclude certain subrepresentations from a representation class [King, 1994]. Fix a stability parameter $\theta \in \mathbb{R}^{Q_0}$ satisfying $\theta \cdot \dim \mathcal{F} = 0$. We use the convention that $\mathcal{F}$ is θ-semistable if every nonzero proper subrepresentation $\mathcal{F}' \subset \mathcal{F}$ satisfies $\theta \cdot \dim \mathcal{F}' \geq 0$. Within the θ-semistable class, choosing $\theta(\mathcal{F}_{\mathrm{triv}}) < 0$ therefore excludes the trivial representation as a subrepresentation, without requiring it to occur as a direct summand. The negative weight alone imposes no such restriction on an arbitrary learned representation. This criterion motivates the moment-map regularizers introduced below, which encourage the balance conditions associated with polystability.

King stability controls subrepresentations directly, whereas the kernel decomposition theorem describes the harmonic contributions of direct summands. Since every direct summand is also a subrepresentation, the stability criterion excludes trivial direct summands as a special case. Together with harmonic injection, these results connect representation structure to the diffusion limit and motivate stability-based geometric control during learning.

Caveats of Degeneration. We emphasize that excluding $\mathcal{F}_{\mathrm{triv}}$ prevents only one specific mode of homological collapse. A learned sheaf may also degenerate into a geometry where $H^0(G;\mathcal{F}) = 0$. In this scenario, all signals decay to zero under continuous linear diffusion with the sheaf geometry held fixed. While θ-stability provides a targeted tool to mathematically exclude the trivial constant-signal mode, it does not universally guarantee that the surviving harmonic space is highly expressive or perfectly aligned with the downstream task.

6 Moment-Map Regularization and Architectural Constraints

The exclusion criterion above requires the learned sheaf to be θ-semistable. Enforcing this definition directly would require controlling all subrepresentations, that is, all families of stalk subspaces compatible with the restriction maps. This does not provide a convenient differentiable training objective. We therefore turn to θ-polystability, a stronger condition that admits a moment-map characterization.

From Semistability to Polystability. A representation is θ-stable if it has total weight zero and every nonzero proper subrepresentation has strictly positive weight, using our sign convention. It is θ-polystable if it is a direct sum of θ-stable representations, each of weight zero. Every θ-polystable representation is therefore θ-semistable. Consequently, polystability is sufficient for our exclusion mechanism: a θ-polystable sheaf with $\theta(\mathcal{F}_{\mathrm{triv}}) < 0$ cannot contain $\mathcal{F}_{\mathrm{triv}}$ as a subrepresentation.

The Moment-Map Criterion. For this discussion, we use the standard complexified setting over $k = \mathbb{C}$. Choose Hermitian inner products on all stalks. The compact gauge group $K_{\mathbf{d}} = \prod_{i \in V \sqcup E} U(d_i)$ then acts unitarily on the representation space. Writing $A_{v,e} := \mathcal{F}_{v \trianglelefteq e}$, we use the moment-map components

$$\mu_v(\mathcal{F}) = -\sum_{e:\, v \trianglelefteq e} A_{v,e}^* A_{v,e}, \qquad \mu_e(\mathcal{F}) = \sum_{v:\, v \trianglelefteq e} A_{v,e} A_{v,e}^*.$$

These signs agree with the semistability convention used above.

King's moment-map criterion [King, 1994, Proposition 6.5], building on the Kempf–Ness correspondence [Kempf and Ness, 1979], states that

$$\mathcal{F} \text{ is } \theta\text{-polystable} \quad \Longleftrightarrow \quad \exists\, g \in G_{\mathbf{d}} :\ \mu_i(g \cdot \mathcal{F}) = \theta_i I_{d_i} \quad \text{for every } i \in V \sqcup E.$$

Here $G_{\mathbf{d}} = \prod_i GL(d_i, \mathbb{C})$ is the complex gauge group. Equivalently, the balance equations hold for $\mathcal{F}$ after a suitable choice of stalk inner products.[1]

This criterion replaces conditions on subrepresentations by matrix balance equations. We use these equations to construct differentiable penalties on the raw restriction maps, with the training inner products held fixed. Exact balance would imply polystability; a small finite penalty is a surrogate and does not by itself certify polystability or exclusion. The real-valued implementation follows the same matrix formulas, replacing Hermitian adjoints by transposes.

[1] King's algebraic GIT construction uses integral stability weights. Our learning surrogate treats the weights as real-valued parameters in the same balance equations.

Central Moment Regularization (CentMM). The balance equations require each moment-map block to be a scalar multiple of the identity. CentMM encourages this form without prescribing the scalar in advance. For a given block μ_i, its closest scalar matrix in Frobenius norm is $\widehat{\theta}_i I_{d_i}$, where

$$\widehat{\theta}_i = \frac{\operatorname{tr}(\mu_i)}{d_i}.$$

We therefore define

$$\mathcal{R}_{\text{cent}}(\mathcal{F}) = \sum_i \left\| \mu_i(\mathcal{F}) - \widehat{\theta}_i I_{d_i} \right\|_F^2 .$$

Every incidence map contributes opposite traces at its source and target, so $\sum_i \operatorname{tr}(\mu_i) = 0$. The implicit weights consequently satisfy $\sum_i d_i \widehat{\theta}_i = 0$. Thus, CentMM encourages the form of the polystable balance equations while selecting the scalar targets analytically from the current geometry. Its training objective is $\mathcal{L} = \mathcal{L}_{\text{task}} + \lambda_\mu \mathcal{R}_{\text{cent}}$. It acts as a geometric balance bias, without imposing a prescribed negative weight on the trivial representation.

Adaptive Stability and Implementation (ThetaMM). ThetaMM instead retains an explicit learnable stability weight θ_i for each graph object. The quadratic balance residual is

$$\mathcal{R}_{\theta\text{-}\mu}(\mathcal{F}; \theta) = \sum_i \|\mu_i(\mathcal{F}) - \theta_i I_{d_i}\|_F^2 .$$

For this quadratic residual alone, CentMM is precisely its analytic minimization over admissible θ:

$$\mathcal{R}_{\text{cent}}(\mathcal{F}) = \min_{\theta \cdot \mathbf{d} = 0} \mathcal{R}_{\theta\text{-}\mu}(\mathcal{F}; \theta).$$

Keeping θ explicit allows us to introduce additional constraints on its weights. This analytic equivalence does not generally extend to the Huber variant or to objectives with an additional θ-norm penalty.

We enforce admissibility by projecting unconstrained parameters $\widetilde{\theta}$ onto the orthogonal complement of the dimension vector:

$$\theta = \widetilde{\theta} - \frac{\widetilde{\theta} \cdot \mathbf{d}}{\|\mathbf{d}\|_2^2} \mathbf{d}.$$

This projection is differentiable and ensures $\theta \cdot \mathbf{d} = 0$. In the negative-weight experiments, we additionally enforce $\theta(\mathcal{F}_{\text{triv}}) \leq -\varepsilon$ for a fixed $\varepsilon > 0$ after each optimization step. The experimental implementation uses averaged residuals, including a Huber variant for ThetaMM and an additional θ-norm penalty, as specified in Appendix A.

Architectural Constraint. If all stalks have dimension d, admissibility gives $0 = \theta \cdot \mathbf{d} = d\,\theta(\mathcal{F}_{\text{triv}})$. Thus, equal-stalk architectures place the trivial representation on a stability wall. Nonuniform dimensions remove this proportionality obstruction, making negative weights admissible in the rectangular $2 \to 3$ and $3 \to 4$ settings studied here.

Exact polystability has a further consequence in these settings. For uniform vertex and edge dimensions $p < q$, every global section is supported on vertices whose incident restriction maps all vanish. Consequently, $H^0(G; \mathcal{F}) = 0$ if each vertex has a nonzero incident map. Indeed, in balancing inner products, a section induces a subrepresentation supported on vertex and edge sets S and T. Semistability and the traced balance equations yield

$$\frac{q}{p} \sum_{e \in T} \theta_e \leq -\sum_{v \in S} \theta_v \leq \sum_{e \in T} \theta_e.$$

Since $\theta_e \geq 0$ and $q > p$, both sums vanish. Hence $\theta_v = 0$ on S, and the vertex balance equations force all incident maps there to vanish. This characterizes the exact balance target and its fixed linear diffusion limit; finite moment penalties remain soft geometric biases during nonlinear NSD training.

7 Empirical Study of Learned Sheaf Geometry

Our experiments investigate the proposed connection between representation structure, stability and geometric control during learning. We first examine the architectural obstruction, then study controlled trivial-component formation and recovery from collapsed representations, and finally investigate how regularization strength influences native NSD geometries. Classification comparisons

use paired initializations and CE-only controls; validation-selected checkpoints and fixed training endpoints are evaluated separately.

The native studies use feature-conditioned NSD architectures based on Bodnar et al. [2022]. Recovery experiments use 16 NSD layers with MLP restriction learners and $3 \to 4$ stalks, while the regularization-strength and learner-capacity studies use $2 \to 3$ stalks. The reported WebKB classification comparisons use one fixed public split per graph and three seeds per condition; their CE-only baselines use neither dropout nor weight decay. Training and diagnostic details are in Appendix A; code, registered configurations and reproduction instructions are available at `https://github.com/ArifDoenmez/NSD-GIT`.

Rectangular Stalks Enable Stability-Wall Access. The equal-stalk obstruction has a direct numerical counterpart. Starting from nonzero admissible weights on the wall $\theta(\mathcal{F}_{\mathrm{triv}}) = 0$, square $3 \to 3$ models remain on the wall in all three seeds, whereas rectangular $3 \to 2$ models leave it in all three. The earlier rectangular controls likewise yield nonzero weights at all ten selected checkpoints, with positive signs under our convention. These observations illustrate the architectural accessibility of nonzero weights.

Negative weights are also compatible with unconstrained sign optimization: six auxiliary $2 \to 3$ runs on Texas and Wisconsin have negative weights at their validation-selected checkpoints. These runs already initialize negatively, so this is not evidence of spontaneous crossing into the negative region. Subsequent negative-ThetaMM experiments explicitly maintain the sign condition. Architectural accessibility and enforced negativity are therefore distinguished throughout the analysis.

Trivial-Component Formation under Dirichlet Minimization. With node signals fixed and total map norm constrained, direct optimization of the incidence matrices produces verified trivial direct summands in all ten synthetic cases and all six Texas/Wisconsin cases. Explicit inclusion and retraction witnesses establish splitting. These controlled matrix-optimization tests complement the native feature-conditioned classification experiments below.

Recovery under Increased Task Load. We next initialize a native NSD model with a verified trivial direct summand and identical hidden node signals inside that copy. The encoder remains trainable and retains access to informative raw features. This creates a controlled collapsed starting representation from which both regularized and unregularized models can recover.

For each seed, we hold a 240-node, 4-regular synthetic graph and its training/validation/test split fixed while increasing the task load from three to six and twelve classes. The easier settings show near-perfect classification and no consistent benefit from regularization. At twelve classes, the distinction becomes pronounced: final mean test accuracy is 29.44% for CE-only training, compared with 65.00% and 76.67% for ThetaMM weights one and ten, respectively. All models achieve 100% training accuracy at this endpoint. The improved test accuracy is accompanied by moment-map blocks that are substantially closer to scalar multiples of the identity, even after accounting for their overall scale (relative central imbalance: 0.80–0.82 without regularization versus 0.16–0.22 with regularization).

This supports a geometric regularization effect on generalization under increased task load. It does not isolate class count as the sole cause, since the fixed sample budget also reduces examples per class. In a separate three-class experiment, an input-ablated control remains at chance, confirming the role of informative inputs in recovery in that setting.

Recovery on WebKB Graphs. We adapt the controlled collapsed initialization to Texas and Wisconsin while retaining their original features, labels and data splits. Both datasets have five classes, compared with twelve in the most demanding synthetic setting. These experiments therefore examine recovery on real heterophilic graphs at a lower class count. At the fixed final endpoint, all models achieve 100% training accuracy, while ThetaMM with weight ten improves test accuracy over its paired CE-only baseline in all six runs. Table 1 reports both validation-selected and final results.

The generalization benefits remain setting- and metric-dependent: the validation-selected Wisconsin accuracy gains, for example, are not accompanied by improved test macro-recall. Moreover, these task differences do not establish removal of a current trivial component as their cause. They demonstrate that the proposed geometric intervention can influence learning and generalization from a known collapsed starting configuration.

Setting	Endpoint	CE-only	ThetaMM, $\lambda_\mu = 1$	ThetaMM, $\lambda_\mu = 10$
Synthetic, 12 classes	Selected	28.89 ± 5.85	60.56 ± 15.49	74.44 ± 1.92
	Final	29.44 ± 8.22	65.00 ± 7.26	76.67 ± 5.00
Texas	Selected	74.77 ± 1.56	77.48 ± 1.56	83.78 ± 4.68
	Final	68.47 ± 1.56	79.28 ± 1.56	77.48 ± 7.80
Wisconsin	Selected	58.82 ± 3.40	60.13 ± 4.08	60.78 ± 1.96
	Final	56.21 ± 5.99	62.75 ± 5.19	69.28 ± 2.26

Table 1: Recovery from controlled collapsed initializations: test accuracy (%, mean $\pm$ sample standard deviation across three paired seeds). Selected checkpoints minimize validation cross-entropy; final checkpoints use 2,000 updates. All final models achieve 100% training accuracy.

Regularization Strength Steers Learned Geometry. Finally, we compare CentMM and ThetaMM at weights $\lambda_\mu \in \{0.1, 1, 10\}$ in native feature-conditioned NSD with 16 and 128 layers. CentMM improves relative central balance at every tested weight in the 16-layer setting. At weight ten, the final relative central residual decreases from 0.547 to 0.218 on Texas and from 0.626 to 0.249 on Wisconsin. The dependence on weight is not uniformly monotone.

Regularization also changes the energy trajectory in a depth-dependent manner. In the affine-plus-$\tanh$ learner study, final energy is lower in all 36 regularized arm–seed contrasts at 16 layers, but higher in 30 of 36 contrasts at 128 layers. These contrasts share their CE-only controls. Thus, in the deeper setting, regularization predominantly produces higher energies than unregularized training, rather than uniformly reducing energy.

The findings show that moment-map regularization steers balance and energy during classification training, with the energy response depending on depth and learner setting. Since both the geometry and node signals evolve, neither higher nor lower energy alone establishes improved information preservation. The empirical evidence supports geometric control during learning, while the exclusion interpretation remains conditional on the stability conditions developed above.

8 Related Work

Neural Sheaf Architectures. Spectral and energy-based analyses relate oversmoothing to feature concentration under repeated propagation [Li et al., 2018, Chen et al., 2020]. Cellular sheaves enrich the diffusion geometry through local restriction maps [Hansen and Ghrist, 2019, Bodnar et al., 2022]; attention and connection-Laplacian variants develop this further [Barbero et al., 2022b,a]. Bodnar et al.'s NSD provides our controlled reference for isolating stalk dimensions and regularization.

Recent methods intervene at different levels: sheaf positional encodings combine structural and semantic information [He et al., 2024], Bayesian networks learn distributions over sheaves [Gillespie et al., 2024], nonlinear diffusion learns the interaction inside the sheaf Laplacian [Zaghen et al., 2024], and joint diffusion co-evolves signals and restriction maps [Hernandez Caralt et al., 2024]. Polynomial NSD learns spectral filters [Borgi et al., 2025], while cooperative sheaves enable directed communication [Ribeiro et al., 2026]. These approaches modify encoding, map learning or propagation; we study representation structure and regularize the balance of the restriction maps.

Quiver Structure and Harmonic Information. Sumray et al. [2024] develop quiver Laplacians and compatible sections, explicitly relating sheaves to subdivided-quiver representations. Hume and Liò [2025] give a categorical account of higher-order message passing and use direct sums of sheaves and harmonic spaces for separation results. These provide related mathematical foundations for our decomposition-based perspective. Bundle Neural Networks use diffusion on flat vector bundles [Bamberger et al., 2025]. Their construction highlights the role of local coordinates: algebraic triviality must be distinguished from homogenization of node features in their original coordinates. Accordingly, our degeneration model also depends on the embedding, signals and task.

Recent Analyses of Sheaf Diffusion. Identity-sheaf baselines achieve competitive results in finite trained networks [Hernandez Caralt et al., 2026], emphasizing the gap between idealized heat-flow limits and complete architectures. Deep NSD changes the propagation operator, normalization and gating to sustain depth-dependent signals [Bourgerie et al., 2026]. Subsequent preprints study relative harmonic information through an index-theoretic criterion [Dong et al., 2026] and local operator

structure through a Bochner–Weitzenböck decomposition [Guadagnuolo et al., 2026]. The latter also studies learned-geometry trajectories. These perspectives complement our subrepresentation-based analysis; neither kernel dimension nor energy alone certifies preservation of discriminative information.

Implicit Bias, GIT and Geometric Control. Implicit-bias results motivate investigating simple learned structures [Rahaman et al., 2019, Gunasekar et al., 2017, Shah et al., 2020, Arora et al., 2019], without establishing a universal NSD optimization law. King stability and the moment-map characterization of polystability are established GIT tools [King, 1994, Kempf and Ness, 1979, Mumford et al., 1994, Kirwan, 1984]. Our contribution connects these tools to learned NSD geometries: representation decompositions motivate a targeted exclusion criterion, the equal-stalk obstruction identifies an architectural limitation, and differentiable balance surrogates provide mechanisms for geometric control during learning.

9 Limitations

While this work provides a representation-geometric framework for studying oversmoothing, we view representation degeneracy as a conceptual perspective rather than a universal explanation. Our decomposition results concern fixed linear sheaf diffusion and motivate investigating the structures learned by feature-conditioned NSD. Stability-based exclusion addresses one specific trivial-component mechanism; the usefulness of the remaining geometry also depends on the learned signals and the downstream task. The equal-stalk obstruction is an algebraic result, while our experiments illustrate its architectural implications and the effects of geometric regularization during learning. The controlled synthetic and WebKB studies support moment-map penalties as structured, setting-dependent geometric biases; broader datasets and architectural comparisons remain directions for future work. Finally, our differentiable penalties encourage balanced moment-map configurations but do not enforce exact polystability during gradient descent. For the rectangular $p < q$ architectures, exact polystability forces the global-section space to vanish when every vertex has a nonzero incident map, so fixed linear diffusion then converges to zero at infinite time. This motivates exploring intermediate regularization strengths that balance geometric control with sufficient freedom for task-adapted learning. The associated exclusion interpretation remains conditional on the corresponding stability conditions.

10 Conclusion

We connect learned NSD geometries with Geometric Invariant Theory through their interpretation as incidence-quiver representations. Relating representation decompositions to the harmonic space of fixed linear diffusion motivates studying trivial or low-complexity components as potential degeneration modes. King stability supplies a conditional exclusion criterion, and the moment-map characterization of polystability motivates differentiable balance regularizers. We identify the exact equal-stalk obstruction to assigning a destabilizing weight to the trivial representation and show how nonuniform stalk dimensions remove it.

Controlled energy minimization reproduces trivial-component formation, while native NSD studies show that regularization steers balance and energy during learning, with depth-dependent responses and setting-dependent generalization benefits in controlled recovery tasks. These findings illustrate geometric control without establishing universal exclusion or an anti-oversmoothing guarantee. The main contribution is a framework connecting representation structure, stability and trainable geometric interventions, opening further study of how learned sheaf geometries support task-relevant information.

Acknowledgements and Disclosures

Arif Dönmez thanks his wife and children for their support and inspiration.

Funding. This work was supported through the authors' affiliated institutions. No additional project-specific funding was received.

Competing interests. The authors are affiliated with DNTOX GmbH. They declare no other competing financial interests.

References

Sanjeev Arora, Nadav Cohen, Wei Hu, and Yuping Luo. Implicit regularization in deep matrix factorization. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Ibrahim Assem, Daniel Simson, and Andrzej Skowroński. *Elements of the representation theory of associative algebras: Volume 1: Techniques of representation theory*, volume 65. Cambridge University Press, 2006.

Jacob Bamberger, Federico Barbero, Xiaowen Dong, and Michael Bronstein. Bundle neural network for message diffusion on graphs. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/2ac879d1865475a7abc8dfc7a9c15c27-Abstract-Conference.html.

Federico Barbero, Cristian Bodnar, Haitz Sáez de Ocáriz Borde, Michael M. Bronstein, Petar Veličković, and Pietro Liò. Sheaf neural networks with connection laplacians. In *Topological, Algebraic and Geometric Learning Workshops 2022*, volume 196 of *Proceedings of Machine Learning Research*, pages 28–36. PMLR, 2022a.

Federico Barbero, Cristian Bodnar, Haitz Sáez de Ocáriz Borde, and Pietro Liò. Sheaf attention networks. In *NeurIPS 2022 Workshop on Symmetry and Geometry in Neural Representations*, 2022b.

Cristian Bodnar, Francesco Di Giovanni, Benjamin Paul Chamberlain, Pietro Liò, and Michael M. Bronstein. Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in GNNs. In *Advances in Neural Information Processing Systems*, volume 35, pages 18527–18541, 2022.

Klaus Bongartz. Some geometric aspects of representation theory. *Algebras and modules, I (Trondheim, 1996)*, 23:1–27, 1998.

Alessio Borgi, Fabrizio Silvestri, and Pietro Liò. Polynomial neural sheaf diffusion: A spectral filtering approach on cellular sheaves. arXiv preprint arXiv:2512.00242, 2025. URL https://arxiv.org/abs/2512.00242v3. Version 3, 2026.

Rémi Bourgerie, Šarūnas Girdzijauskas, and Viktoria Fodor. Deep neural sheaf diffusion. arXiv preprint arXiv:2605.19021, 2026. URL https://arxiv.org/abs/2605.19021v3. Version 3.

Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou, and Xu Sun. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3438–3445, 2020.

Justin Michael Curry. *Sheaves, Cosheaves and Applications*. PhD thesis, University of Pennsylvania, 2014.

Harm Derksen and Jerzy Weyman. Quiver representations. *Notices of the American Mathematical Society*, 52(2):200–206, 2005.

Junwen Dong, Yuhan Peng, Hao Li, Huitao Feng, and Kelin Xia. Demystifying oversmoothing in sheaf neural networks: An index-theoretic criterion. arXiv preprint arXiv:2608.16180, 2026. URL https://arxiv.org/abs/2608.16180.

Arif Dönmez. ArchTrial, 2026. URL https://github.com/ArifDoenmez/archtrial. Software, version 0.1.0; commit 8210f084ee5f.

Patrick Gillespie, Layal Bou Hamdan, Ioannis Schizas, David L. Boothe, and Vasileios Maroulas. Bayesian sheaf neural networks. arXiv preprint arXiv:2410.09590, 2024. URL https://arxiv.org/abs/2410.09590v2. Version 2, 2025.

Francesco Maria Guadagnuolo, Alessio Borgi, Dorina Thanou, and Pietro Liò. A Bochner–Weitzenböck decomposition for cellular sheaf laplacians, curvature and neural sheaf diffusion. arXiv preprint arXiv:2610.28455, 2026. URL https://arxiv.org/abs/2610.28455.

Suriya Gunasekar, Blake E Woodworth, Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. Implicit regularization in matrix factorization. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

Jakob Hansen and Robert Ghrist. Toward a spectral theory of cellular sheaves. *Journal of Applied and Computational Topology*, 3:315–358, 2019.

Yu He, Cristian Bodnar, and Pietro Liò. Sheaf-based positional encodings for graph neural networks. In *Proceedings of the 2nd NeurIPS Workshop on Symmetry and Geometry in Neural Representations*, volume 228 of *Proceedings of Machine Learning Research*, pages 1–18. PMLR, 2024. URL https://proceedings.mlr.press/v228/he24a.html.

Ferran Hernandez Caralt, Guillermo Bernárdez Gil, Iulia Duta, Pietro Liò, and Eduard Alarcón Cot. Joint diffusion processes as an inductive bias in sheaf neural networks. In *Proceedings of the Geometry-grounded Representation Learning and Generative Modeling Workshop (GRaM)*, volume 251 of *Proceedings of Machine Learning Research*, pages 249–263. PMLR, 2024. URL https://proceedings.mlr.press/v251/hernandez-caralt24a.html.

Ferran Hernandez Caralt, Mar Gonzàlez i Català, Adrián Bazaga, and Pietro Liò. On the necessity of learnable sheaf laplacians. arXiv preprint arXiv:2603.05395, 2026. URL https://arxiv.org/abs/2603.05395.

Jacob Hume and Pietro Liò. On the sheafification of higher-order message passing. arXiv preprint arXiv:2509.23020, 2025. URL https://arxiv.org/abs/2509.23020.

George R. Kempf and Linda Ness. The length of vectors in representation spaces. In *Algebraic Geometry*, pages 233–243. Springer, 1979.

Alastair D. King. Moduli of representations of finite dimensional algebras. *The Quarterly Journal of Mathematics*, 45(4):515–530, 1994. doi: 10.1093/qmath/45.4.515.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Frances Clare Kirwan. *Cohomology of Quotients in Symplectic and Algebraic Geometry*. Princeton University Press, 1984.

Hanspeter Kraft. Geometric methods in representation theory. In Maurice Auslander and Emilio Lluis, editors, *Representations of Algebras*, volume 944 of *Lecture Notes in Mathematics*, pages 180–258. Springer, 1982. doi: 10.1007/BFb0094059.

Hanspeter Kraft. *Geometrische Methoden in der Invariantentheorie*. Aspekte der Mathematik. Vieweg+Teubner, 2 edition, 1985. doi: 10.1007/978-3-663-10143-7.

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*, 2018.

David Mumford, John Fogarty, and Frances Kirwan. *Geometric Invariant Theory*. Springer, 3 edition, 1994.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. In *International Conference on Learning Representations*, 2020.

Vladimir L Popov and Ernest B Vinberg. Invariant theory. In *Algebraic Geometry IV: Linear Algebraic Groups Invariant Theory*, pages 123–278. Springer, 1994.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning*, pages 5301–5310. PMLR, 2019.

Markus Reineke. Moduli of representations of quivers, 2008. URL https://arxiv.org/abs/0802.2147.

André Ribeiro, Ana Luiza Tenorio, Juan Belieni, Amauri Souza, and Diego Mesquita. Cooperative sheaf neural networks. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/ea1a7f7bc0fc14142106a84c94c826d0-Abstract-Conference.html.

Ralf Schiffler. *Quiver Representations*. Springer, 2014.

Harshay Shah, Kaustubh Tamuly, Aditi Raghunathan, Prateek Jain, and Praneeth Netrapalli. The pitfalls of simplicity bias in neural networks. In *Advances in Neural Information Processing Systems*, volume 33, pages 9573–9585, 2020.

Allen Dudley Shepard. *A cellular description of the derived category of a stratified space*. PhD thesis, Brown University, 1985.

Otto Sumray, Heather A. Harrington, and Vidit Nanda. Quiver laplacians and feature selection. arXiv preprint arXiv:2404.06993, 2024. URL https://arxiv.org/abs/2404.06993.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

Olga Zaghen, Antonio Longa, Steve Azzolin, Lev Telyatnikov, Andrea Passerini, and Pietro Liò. Sheaf diffusion goes nonlinear: Enhancing GNNs with adaptive sheaf laplacians. In *Proceedings of the Geometry-grounded Representation Learning and Generative Modeling Workshop (GRaM)*, volume 251 of *Proceedings of Machine Learning Research*, pages 264–276. PMLR, 2024. URL https://proceedings.mlr.press/v251/zaghen24a.html.

A Experimental Details

Data and comparisons. The current WebKB studies use frozen Texas and Wisconsin split-0 inputs from Geom-GCN [Pei et al., 2020]. Features are row-normalized nonnegative bag-of-words vectors. Removing self-loops and duplicate edge directions gives 183 nodes/279 undirected edges for Texas and 251/450 for Wisconsin, with 1,703 features and five classes each. Actual train/validation/test counts are 87/59/37 and 120/80/51. Masks are retained unchanged; one Texas class is absent from its training and test masks. Macro-recall averages over classes present in the evaluated mask. Features and graph structure are available transductively; task gradients use training labels only.

Study	Stalks	Layers	Updates	Optimizer / LR	Seeds
Common-wall pilot	$3 \to 3, 3 \to 2$	4	≤ 1500	Adam / 10^{-2}	3
Free-sign pilot	$2 \to 3$	linear diffusion	600	Adam / 10^{-2}	3/graph
Dirichlet, K_4	$2 \to 3$	–	5000	SGD / 0.005	10
Dirichlet, WebKB	$2 \to 3$	–	5000	SGD / 0.005	3/graph
Native weight study	$2 \to 3$	16, 128	2000	Adam / 10^{-4}	3/graph
Recovery studies	$3 \to 4$	16	2000	Adam / 10^{-3}	3/setting

Table 2: Protocols underlying the main empirical results. Recovery includes the initial three-class control, the class-load study and WebKB adaptation. Fixed training endpoints and validation-selected checkpoints are distinguished throughout.

Native architecture and objectives. We use a local feature-conditioned General-NSD implementation based on Bodnar et al. [2022], with 20 channels, learned stalk/channel mixing, residual coefficients and propagation ELU. Each layer has its own restriction learner, shared across incidences. The weight study uses an ordered-endpoint linear learner followed by $\tanh$; recovery uses a $120 \to 128 \to 12$ ReLU MLP with output $\tanh$. Each output matrix is capped at Frobenius norm one. Native weight and recovery comparisons have no dropout, weight decay, clipping, schedule or early stopping. The separate common-wall pilot uses dropout 0.7, weight decay 0.005, gradient clipping at one and validation patience 200. The free-sign pilot generates maps once per forward and holds them fixed through linear diffusion; it does not measure nonlinear layer depth.

Native ThetaMM uses Smooth-L1/Huber loss with transition one, averaging residual entries separately over vertex and edge blocks, summing those group means and averaging over layers. Its objectwise θ is shared across layers, with mean-squared norm penalty 10^{-4}. CentMM uses the same averaging with squared residuals and per-layer analytic targets $\widehat{\theta}_i = \operatorname{tr}(\mu_i)/d_i$, without learned θ or a θ-norm penalty. Doses are $\lambda_\mu = 0.1, 1, 10$ in the weight study and one/ten in recovery. Equal weights across Huber and quadratic losses are not matched penalty strengths. Negative-ThetaMM enforces $\theta \cdot \mathbf{d} = 0$ and $\sum_i \theta_i \leq -0.01$ after updates; the free-sign pilot retains only admissibility after its negative initialization.

Controlled energy and collapsed initialization. Direct energy tests optimize free incidence matrices with fixed node signals and total squared map norm one. On K_4, $x_v = e_1$; on WebKB, fixed unit signals are obtained without labels from the input features. These calibrations use neither classification loss nor native end-to-end map learning. Their regularized controls use quadratic moment residuals; exact settings and tolerance-sensitive outcomes are retained in the accompanying protocols.

Recovery instead begins with native maps $A_{v,e} = \frac{1}{2} u v^\top$ for unit $u \in \mathbb{R}^4$, $v \in \mathbb{R}^3$. The vectors $s_v = v$, $s_e = \frac{1}{2} u$ and covectors $p_v = v^\top$, $p_e = 2u^\top$ verify a trivial direct summand, leaving complementary dimensions two and three. The encoder has zero weight and bias $vc^\top$ for 20 positive channel coefficients. All parameters remain trainable, with informative raw inputs retained. On irregular WebKB graphs, channel propagation weights additionally start at zero to preserve identical hidden states. Initial raw and propagated-input energies are zero; normalized hidden energy need not be zero.

Synthetic recovery uses 4-regular circulants with offsets one and seven, randomly relabeled within each seed. The class-load study fixes 240 nodes and 120/60/60 masks across $K = 3, 6, 12$; labels before relabeling are index modulo K. Orthonormal class means have amplitude 1.5 in 12 input dimensions, with Gaussian noise standard deviation 0.5 and train-only centering/scalar-RMS pre-

processing. Training examples per class decrease from 40 to 20 to 10. The separate three-class control uses 120 nodes and 60/30/30 masks and includes a zero-input ablation.

Measurements and numerical verification. For the actual layer input H_ℓ, the reported network energy and common central imbalance are

$$E_\ell = \|\delta_{\mathcal{F}^{(\ell)}} N_\ell H_\ell\|_F^2, \quad (N_\ell)_v = \left(I + \sum_e A_{v,e}^* A_{v,e} \right)^{-1/2}, \quad b_{\mathrm{cent}} = \sqrt{\frac{\sum_i \|\mu_i - \widehat{\theta}_i I\|_F^2}{\sum_i \|\mu_i\|_F^2}}.$$

The sum in N_v is over incidences at v; inverse square roots use 20 Newton–Schulz iterations. Energy omits a factor $1/2$. Summaries take a layer median per model and then a median across seeds. Relative imbalance is distinct from the optimized averaged Huber term. Splitting requires compatible inclusions and retractions, $A_{v,e} s_v = s_e$, $p_e A_{v,e} = p_v$ and $p_i s_i = 1$ on every object. Relative SVD cutoffs are 10^{-10}, 10^{-8} and 10^{-6}, with witness equation errors at most 10^{-8}. Coarser thresholds recover approximate witnesses in two strongly regularized synthetic endpoints; numerical non-detection is not an exact exclusion theorem.

Selection and reproducibility. Dose checkpoints are evaluated initially and every 25 updates; recovery uses every five updates through 100 and every 25 thereafter. Minimum validation CE selects checkpoints, with earliest ties, before cohort test evaluation. Fixed final endpoints remain separately reported. CPU float64 and one PyTorch/BLAS thread per worker are used for the current native and direct-energy studies; the common-wall pilot uses MPS float32. Earlier ten-split weights are archival sign audits, not current classification replications. Studies were fixed before their own execution; follow-ups used knowledge of preceding results. We organize model variants, hypotheses and comparison configurations with ArchTrial [Dönmez, 2026], a command-line tool for structuring architecture experiments. Our project-specific runners additionally record paired initialization checks and checkpoint-selection locks. Full protocols, seed lists and individual endpoints are available in the public code repository linked in Section 7; layers and shared-control contrasts are not independent replications.

B Additional Results for Central Moment Regularization

CentMM provides a geometric baseline that determines its scalar targets analytically rather than learning stability weights. Table 3 summarizes its effect in the verified native weight study. At 16 layers, every tested CentMM weight improves relative central balance against CE-only on both graphs. Increasing the weight is not uniformly monotone: Wisconsin reaches its lowest residual at weight one. At 128 layers, the changes are smaller and not uniformly favorable at lower weights. This complements ThetaMM by showing that geometric control can already arise from enforcing the central form of moment blocks without explicit stability-weight optimization.

Graph	Layers	CE-only	CentMM 0.1	CentMM 1	CentMM 10
Texas	16	0.547	0.254	0.235	0.218
Texas	128	0.585	0.594	0.585	0.560
Wisconsin	16	0.626	0.315	0.227	0.249
Wisconsin	128	0.612	0.589	0.576	0.563

Table 3: Final relative central imbalance b_{cent} in the native weight study. Each entry is a median over three seeds of the per-model layer medians. The numbers after CentMM denote λ_μ. Lower values indicate moment blocks closer to scalar matrices relative to their overall scale.

These measurements concern relative balance, not exact polystability or component removal. The additional registered all-layer numerical exclusion margin was not achieved in the native dose cohort. Complete task endpoints, including unfavorable outcomes and both checkpoint policies, remain in the machine-readable records. The separate learner-capacity and scale audits are also retained there without repeating their full tables or trajectories in this PDF.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction state the connection between learned NSD geometries and GIT, the decomposition and stability-wall arguments, and the empirical illustrations of geometric control. Fixed linear diffusion, conditional exclusion and the scope of the classification evidence are distinguished explicitly.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 9 discusses the fixed-linear setting, surrogate stability conditions, limited datasets and seeds, and absence of a universal accuracy or anti-oversmoothing guarantee. The conclusion retains these qualifications.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The assumptions, functorial decomposition argument, harmonic-injection proof and stability-wall calculation are provided in Sections 4–6. The moment-map characterization used in Section 6 is explicitly attributed to King.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Appendix A specifies the data processing, architectures, objectives, training and checkpoint selection. Code, frozen configurations, seed lists and recorded endpoints are provided at `https://github.com/ArifDoenmez/NSD-GIT`, with commands mapping the main findings to their registered experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: Code, registered configurations, environment specifications and reproduction commands are publicly available at `https://github.com/ArifDoenmez/NSD-GIT`. The release provides checksum-verified acquisition of the original public WebKB inputs, synthetic data generation, training entry points and stored endpoint analyses. Archival rebuttal diagnostics are distinguished from the newer camera-ready studies.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section 7 and Appendix A describe the datasets, fixed splits, seeds, architectures, objectives, optimizers and checkpoint-selection rules. Exact paired configurations and study-specific commands are included in the public repository.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Table 1 reports means and sample standard deviations across three paired seeds. The repository separately retains the original ten-split rebuttal comparisons and their paired-bootstrap intervals, which include zero; we do not infer general predictive superiority from these comparisons.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Appendix A reports devices, numerical precision, thread counts and training budgets for the current studies. The repository provides the environment recipe and per-study protocols; the archival GPU/CPU fallback is documented separately from the current native CPU studies.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

Justification: The research involves theoretical advancements in graph neural networks and benchmarking on standard academic datasets, adhering fully to the NeurIPS Code of Ethics.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [N/A]

Justification: This is a foundational theoretical paper focusing on the optimization geometry of graph neural networks and does not present direct, immediate negative societal impacts or specialized end-user applications.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The paper does not release high-risk models, generative systems, or newly scraped web datasets that require specific safeguards.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The NSD baseline and data sources are credited. The public code repository includes the Apache-2.0 license, attribution notice and upstream data-acquisition references. WebKB inputs are downloaded explicitly from their original source rather than redistributed under the code license.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The released research code includes documented models, registered hypotheses and configurations, reproduction commands, source checksums and limitations. Controlled synthetic inputs are generated by the released scripts; no new standalone benchmark dataset is claimed.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The research methodology is strictly algorithmic and theoretical; it does not involve crowdsourcing or human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The research does not involve human subjects, making IRB approval inapplicable.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: No LLMs were used as a component of the core methods in this research. Any use of LLMs was limited to writing and editing assistance, which does not impact the methodology or scientific contributions of the paper.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.