Supplementary Information for Proteus: Adaptive Scale-Space Analysis for Learning Non-Linear Fuzzy Manifold Memberships

S1. Initial Dimensionality Reduction

Randomized SVD: Sample up to 20 000 rows (or all if small), compute a randomized SVD with sketch width $k_{\text{target}} \approx 256$ and oversampling 20. Determine rank by cumulative energy $\geq 99.9\%$ and/or Levina–Bickel intrinsic dimension; keep $r = \max(r_{0.999}, 2\lceil D_{\text{LB}} \rceil)$ if it yields a $\geq 5\%$ reduction; otherwise skip projection. Optionally append 32 JL vectors to tighten distance preservation.

$$P = V_{1:r}^{\top} \;\Rightarrow\; Z = X\,P\,, \quad Z \in \mathbb{R}^{N \times r}.$$

Rationale: Cuts ANN costs and per-sample updates by ×5–×40 on high-d data while retaining $\geq 99.9\%$ variance and local distances.

S1.1 Optional Pre-processing via Marginal Gaussianisation

Workflow: For datasets where individual features exhibit high skew or heavy tails, an optional pre-processing step can significantly accelerate convergence. For each feature x_j, its empirical CDF $\hat{F}_j$ is computed on a data subsample. The features are then transformed via the inverse error function to be marginally Gaussian: $z_j = \sqrt{2}\,\operatorname{erf}^{-1}(2\hat{F}_j(x_j) - 1)$.

Linearity Test and Model Selection: After this transform, statistical linearity tests (e.g., Graph-Laplacian Curvature Energy and distance correlation) are run on the z-space. If the data is found to be linearly separable to within a given tolerance ($p \geq 0.01$), the expensive fitting of a global `Glow` model can be skipped entirely for that recursion level, and the algorithm can proceed directly to the geometric audit. This provides a principled mechanism for avoiding unnecessary non-linear transformations.

S2. First-Principles Derivations

S2.1 Geometric Grid Spacing

Normalized responses in scale-space obey $\Delta_{\text{FWHM}}(\log\sigma) \approx 0.586$. If two peaks are separated by less than this, they are not reliably distinguishable. Therefore, choose adjacent scales to differ by $\log\sigma_{j+1} - \log\sigma_j = \Delta \approx \frac{1}{2}\Delta_{\text{FWHM}}$, giving $\sigma_{j+1}/\sigma_j = e^{\Delta} \approx \sqrt{2}$. Since $\tau = \sigma^2$, the geometric ratio in τ becomes $r = \tau_{j+1}/\tau_j = (\sigma_{j+1}/\sigma_j)^2 \approx 1/2$. In practice we span both directions, so we use $r = 1/\sqrt{2} \approx 0.71$ in σ (7–9 points/decade) and map consistently into τ.

S2.2 EWMA Weight from Neighborhood Size

EWMA recursion $m_{t+1} = (1-\alpha)m_t + \alpha x_{t+1}$ has half-life $T_{1/2} = \ln(0.5)/\ln(1-\alpha) \approx \ln 2/\alpha$ for small α. Set $T_{1/2}{=}k$ (one neighborhood of updates) $\Rightarrow \alpha = \ln 2/k$. This ties estimator memory to the actual neighborhood support.

S2.3 Dual-Rate Motion from Variance Correction

Under isotropy, variance $\sigma_i^2 \approx \mathbb{E}\|x{-}w_i\|^2$. A radial shift δr changes variance by $\Delta\sigma_i^2 \approx 2\sqrt{\sigma_i^2}\,\delta r$. Spreading over $H{=}k$ effective updates gives

$$\eta_{\mathrm{GNG},i} \approx \frac{\Delta\sigma_i^2}{2H\,\sigma_i^2} = \frac{1}{2k}\Big(1 - \frac{\sigma_i^2}{\tau}\Big)\,; \qquad \eta_{\mathrm{GNG},i} = \frac{\ln 2}{2k}\Big(1 - \frac{\sigma_i^2}{\tau}\Big) \text{ with half-life normalization.}$$

For centering, with $\delta_{\min} = \kappa(1{-}r)\sqrt{\tau}$ and expected drift per update $\approx \sqrt{\tau}$, set $k\,\eta_{\mathrm{cent}}\,\sqrt{\tau} = \delta_{\min}$ $\Rightarrow \eta_{\mathrm{cent}} = \kappa(1{-}r)/k$. Unless stated otherwise, we use a default of $\kappa = 0.5$.

Deferred nudge mechanism: Node positions are updated only when the accumulated nudge $\mathbf{a}_i$ exceeds the resolution threshold $\delta_{\min}$, ensuring mesh changes occur only when meaningfully resolvable at the current scale. This prevents high-frequency noise from destabilizing the mesh while maintaining geometric accuracy.

S2.4 Control-to-Threshold Mapping

Constrain growth to local intrinsic dimension by mapping optimizer control to thresholds: $\tau_{\mathrm{global}} = -D_{\mathrm{subspace}}\,\log(1{-}s_{\mathrm{control}})$ and $\tau_{\mathrm{local},i} = -d_{\mathrm{final},i}\,\log(1{-}s_{\mathrm{control}})$.

S3. Statistical Gauntlets

Unless noted, we adopt significance defaults consistent with Stage 1: Welch's t tests at $p{=}0.05$ (one-tailed where directional) for neighborhood-stability triggers, and Wilson one-sided bounds at 90% (i.e., $z{=}1.2816$) for link significance comparisons.

S3.1 Link Pruning

Power shield: protect a new link until $n_{\mathrm{eff}} \geq n_{\mathrm{req}}$ from an analytic power formula for proportions. *Relative weakness:* compute Wilson one-sided upper bound ν^+ for link proportion p with count n:

$$\nu^+ = \frac{p + z^2/(2n) + z\sqrt{p(1-p)/n + z^2/(4n^2)}}{1 + z^2/n}, \; z{=}1.2816.$$

The Wilson score interval is used in place of a standard t-test as it is an "exact" test for binomial proportions that does not assume normality, making it robust and accurate even for the low-hit-count links common in early training. Vote to prune if $\nu^+ <$ median peer significance; require bilateral agreement from both endpoints.

S3.2 Node Pruning

Trigger: require neighborhood stability (Welch t on $\tilde{\rho}$). *Fair chance:* Poisson arrival-rate test $\lambda_{\mathrm{est}} = h_i/\text{lifetime} \geq 2/\delta_{\min}$. *Relative:* prune if hits$<$ 0.5× neighbor mean *and* variance $\sigma_i^2 < 0.5\,\tau_{\mathrm{local},i}$.

The composite rule for node pruning (requiring both low relative hits *and* low local variance) is a critical safeguard. It ensures the system only removes nodes that are both statistically insignificant (low-mass) and located in regions that have already been sufficiently linearized. This prevents the algorithm from mistakenly punching holes in the manifold in areas that are still curved and under active refinement.

S3.3 Node Pruning (Stage 2, simplex-arbitrated)

Arbitration by simplex star: A deletion proposal for node i is decided by a vote over its incident simplexes $\mathcal{S}(i)$; require majority agreement based on local stability and activity metrics. *Geometric veto:* Always perform a reconstruction error test; if removal of i causes a significant increase in local reconstruction error (tent-pole behavior), veto the deletion. *Rationale:* Stage 2's geometry is mediated by the simplicial complex; arbitration by $\mathcal{S}(i)$ preserves structural integrity in complex regions.

S4. Torsion and Shape Quality

The geometric audit operates on the principle that error responses depend on local rigidity: in *plastic* regions, error produces geometric adjustments; in *rigid* regions, error accumulates as structural stress requiring topological resolution. Let $\bar{\mathbf{w}}$ and $\bar{\mathbf{m}}$ be barycenters. With $E = [\mathbf{w}_{v_1}-\bar{\mathbf{w}} \ \cdots \ \mathbf{w}_{v_d}-\bar{\mathbf{w}}]$ and $M = [\mathbf{m}_{v_1}-\bar{\mathbf{m}} \ \cdots \ \mathbf{m}_{v_d}-\bar{\mathbf{m}}]$, define

$$\Omega_S = M^\top E - E^\top M\,, \qquad \kappa_S = \|\Omega_S\|_F\,; \qquad R_S = \kappa_S/\tau^*.$$

Torsion Ladder: *Ignore* if $R_S<0.05$ (numerical noise). *Keep as-is* if $0.05 \leq R_S<0.30$ (tolerable curvature). *Geometric fix* if $0.30 \leq R_S<0.60$: perform a torsion-aligned split with explicit placement below. *Non-linear upgrade* if $R_S \geq 0.60$: attach a local mini-NSF (see S7) to the affected patch. If the fitted mini-NSF fails to reduce R_S below 0.60 on the patch, perform a torsion-aligned split as a fallback.

A split is only permitted if the parent simplex (or node) has accumulated sufficient hits to pass a budget check ($h_{\text{parent}} \geq 2\cdot h_{\text{prune}}$), ensuring that the resulting children are not immediately vulnerable to pruning under the framework's statistical gauntlets.

Split placement: Insert the new node at the midpoint of the simplex's longest edge projected onto the principal axis of Ω_S (dominant torsion axis). If any child has $Q_S<0.25$, redo via centroid star split.

Shape quality safeguard: The simplex shape metric $Q_S = d\,r_{\text{in}}/R_{\text{circ}} \in (0,1]$ uses the inradius r_{in} and circumradius R_{circ}. Poor quality ($Q_S < 0.25$) indicates numerical instability risk; the centroid split fallback ensures stable face normals and well-conditioned gradients.

S5. Dual Flow Mechanisms

Accumulate facet pressures for winning simplex S via $\Delta p_f \propto \max(0, (\mathbf{x}-\bar{\mathbf{w}}_S)^\top \mathbf{n}_f)$. This *fractional tallying* avoids winner–take–all artifacts and yields smoother, more stable boundary pressure estimates than single-face voting. On the dual graph, solve

$$\min_{\{p_f\}} \ \lambda \sum_f (p_f-\hat{p}_f)^2 + \mu \sum_S \|A_S\,\mathbf{p}_S\|_2^2\,,$$

with Gauss–Seidel iterations $p_f \leftarrow (\lambda\hat{p}_f + \mu \sum_{S\ni f}(A_S^\top A_S\,\mathbf{p}_S)_f)/(\lambda + \mu\ \deg(f))$, or equivalently by loopy BP messages until convergence.

S6. Masking for Missing Modalities

For input $\mathbf{x} \in \mathbb{R}^d$ with structured sparsity, build boolean mask $\mathbf{m} = (x \neq 0)$. Compute distances and residuals only on active dimensions: $\|\mathbf{x}-\mathbf{w}\|^2 = \sum_{j:m_j}(x_j-w_j)^2$; residual $\mathbf{e} = \mathbf{x}-\mathbf{w}$ with e_j=0 if m_j=0. Apply the mask to all vector updates (moments, nudges, Oja), and accumulate variance only on active dimensions. Optionally normalize updates by the active count to equalize per-sample influence. This preserves geometry and statistics in multi-modal fusion where zeros indicate missingness rather than small values.

S7. Non-Linear Warp Strategies

When geometric refinement (node moves and splits) is insufficient to resolve high residual curvature, Proteus can apply a non-linear warp. The core strategy is a hybrid approach that adapts to the nature of the curvature, measured by the torsion coverage P_κ (the fraction of simplices with torsion ratio $R_S \geq 0.30$).

Global vs. Patch-wise Strategy:

- If curvature is **widespread ($P_\kappa > 50\%$)**, a shallow *global Glow* is trained. Its invertible 1×1 convolutions mix dimensions efficiently, providing a good baseline correction for system-wide bends.
- If curvature is **patchy ($P_\kappa \leq 25\%$)**, compact and expressive *mini-Normalizing Spline Flows (NSFs)* are attached only to the highest-torsion patches. This avoids wasting model capacity on regions that are already linear.
- If curvature is **mixed ($25\% < P_\kappa \leq 50\%$)**, a moderate global Glow is combined with mini-NSFs on the worst ∼10% of patches.

Patch Identification and Mini-NSF Training: A facet-adjacency graph is built over all "hot" simplices ($R_S \geq 0.60$). Leiden community detection finds contiguous patches of high torsion, with resolution auto-tuned to ensure no patch exceeds a size cap $P_{\max}$ (e.g., 64 simplices). Each mini-NSF is a compact model (2-3 layers of rational-quadratic spline couplings, 64-128 hidden units, 8-12 bins) trained for a small number of epochs on its routed data, with the global model frozen. If, after training, any simplex in the patch still exhibits $R_S \geq 0.60$, attempt a torsion-aligned split as a geometric fallback.

S8. Numerical Stability and Implementation

Warm-start and state reset: On entering Stage 2 at fixed τ^*, shrink EWMA moments by a decay factor $\gamma_0 \approx 0.7$, carry hit/link counts unchanged, and reset the nudge accumulator $\mathbf{a}$ to zero to ensure stable dynamics under the richer neighborhood model ($k \approx d$). **Conditioning for Ω_S:** Use QR (preferred) or thin SVD to stabilize E; scale residuals M by a local norm to avoid magnitude blow-up; compute $\Omega_S = M^T E - E^T M$ from stabilized factors; cache per-simplex factors and recompute only after geometric moves/splits.

Small constants and clipping: Set ε in ρ_i to $10^{-8}-10^{-6}$ depending on data scale; clip η_{GNG} to $|\eta| \leq 0.3$ (Stage 1) and ≤ 0.05 (Stage 2); constrain spline slopes/bins to avoid near-singular Jacobians; bound $\|\mathbf{a}_i\|$ growth between apply steps.

Cache face normals: Precompute and cache per–simplex face normals; recompute only after deferred moves or splits affecting a vertex. This keeps torsion and face-pressure updates $\mathcal{O}(d^2)$ and stable.

ANN configuration (HNSW): Typical defaults: M=16−32, efConstruction=100−400, efSearch=50−200; increase efSearch modestly near convergence to improve neighbor stability. Persist indices per recursion level; rebuild only after significant mesh mutation.

S8.1 Practical Implementation Notes

Rank-1 Oja vs. Scalar Variance. While the core theory holds with a simple scalar variance tracker, the specified implementation uses a rank-1 Oja's rule update to track the dominant principal component of the local error. This provides a more accurate, data-driven vector for placing new nodes during a split, at the cost of one extra stored vector per node. For ambient dimensions $d \lesssim 10,000$, the improved split quality justifies the modest increase in memory, and this method is recommended.

Initial Simplex Seeding. Upon entering Stage 2, the initial simplicial complex is seeded from the Stage 1 node-edge graph using a fast, "Greedy Chaining" heuristic. This heuristic rapidly discovers the majority of d-simplices, providing a robust starting point for the dynamic refinement process.

S8.2 Cost Envelopes

Stage 1 per-epoch time is $\mathcal{O}\big(N\,d\,k\,\log N_{\text{nodes}}\big)$ with HNSW queries; space is $\mathcal{O}(N_{\text{nodes}}\,d + N_{\text{nodes}}\,\bar{d})$. Stage 2 initial simplex seeding by Greedy Chaining is $\mathcal{O}\big(N_c\,d_{\text{cand}}^2\,d_c\big)$; torsion audits are $\mathcal{O}\big(N_{\text{simplices}}\,(d{+}1)^2\,d\big)$. The dual graph used by the belief propagation solver has size $\mathcal{O}(N_{\text{simplices}}\,d)$.

S9. Robustness and Edge Cases

Outliers: Use Huber-like weighting on residuals in moment updates or winsorize top 0.5% residual norms; optionally cap per-sample hit contributions.

Boundaries and orientation: For open meshes, treat boundary faces with Neumann-like conditions in the dual flow (no flux across exterior). In non-orientable regions, operate on local orientation patches to keep face normals consistent.

Sparse patches: Require a minimum routed sample count $n_{\min} = \max(10\,d, 1000)$ before fitting a mini-NSF; otherwise defer to geometric refinement.

S10. Evaluation Protocols

Defaults: Stage 1 $k{=}8$, $\alpha{=}\ln 2/8$, grid $r{=}1/\sqrt{2}$; Stage 2 $k{\approx}d$, $\alpha{=}\ln 2/k$. Stopping: CV<0.01 and $R_S{<}0.30$. PH: VR complex, max dim 2, filtration up to $1.5\,\sigma_\star$. MMD: RBF kernel, bandwidth median heuristic, 1 000 samples/1 000 tests. Ablations: remove Stage 1 scaffold, Torsion Ladder, or Dual Flow.

Persistent Homology: Vietoris–Rips (max dim 2), filtration to $1.5\,\sigma_\star$; report bottleneck/2-Wasserstein. **MMD:** RBF kernel (median heuristic), 1000 samples/1000 tests, 5 seeds. **Log-likelihood:** Sum of face-pressure potentials within simplex plus normalization; use held-out data.

S11. Algorithmic Details

The following procedures reference node and link bookkeeping commonly used in the implementation: nodes track an *update count* and creation index N_{creation} in addition to positions and EWMA statistics; links maintain directed counters and recent activity snapshots. These fields support pruning power analyses and tenure checks but are not essential for the mathematical description.

Flow-conserving initialization after a split. Let C be the parent simplex with vertices $V(C) = \{v_1, \ldots, v_{d+1}\}$ and let $h_{\text{old}}(v_i, v_j)$ denote pre-split internal hit counts between vertices. When inserting a new vertex k, initialize the new tallies by

$$h_{\text{new}}(k, v_i) \;=\; \tfrac{1}{2} \sum_{\substack{v_j \in V(C) \\ j \neq i}} h_{\text{old}}(v_i, v_j)\,.$$

This preserves the total internal mass $H_{\text{total}}(C)$ and approximately preserves the local flow gradient.

Data structures (Stage 2 additions). Each simplex maintains small fields to support the torsion audit and dual-flow:

- `face_pressures`: array of size $d+1$ for facet pressures
- `torsion_sum`, `torsion_sum_sq`: accumulators for κ_S
- `simplex_update_count`: number of updates contributing to the audit

Algorithm 1 Stage 1 at Fixed τ

```
procedure RunToEquilibrium($X, \tau$)
    Initialize mesh $\mathcal{M} \leftarrow \varnothing$    ▷ Create links between BMU$_1$ and BMU$_2$ on first co-activation; maintain directed weighted counters $C_{12}, C_{21}$
    repeat
        Sample $\mathbf{x}$ from $X$
        $\mathcal{N}_k \leftarrow$ QueryANN($\mathbf{x}, k$)    ▷ Find $k$ nearest nodes
        ApplyRankWeightedUpdates($\mathbf{x}$, $\mathcal{M}$, $\mathcal{N}_k$)
        for each updated node $i$ do
            if $\|\mathbf{node}_i.\mathbf{a}\| \geq \delta_{min}$ then    ▷ Deferred move
                $\mathbf{node}_i.\mathbf{w}$ += $\mathbf{node}_i.\mathbf{a}$; $\mathbf{node}_i.\mathbf{a} \leftarrow 0$; RefreshANN($\mathbf{node}_i$)
            end if
            if $\mathbf{node}_i.\sigma^2 > \tau_{local,i}$ then    ▷ Split rule
                SplitNode($\mathbf{node}_i$)    ▷ Flush nudge; place child at $\mathbf{w}_i+\mathbf{m}_i$ along $\mathbf{u}_i$; shrink-inherit stats
            end if
        end for
        ApplyPruningGauntlets($\mathcal{M}$)    ▷ Per SI S3
    until ComputeCV($\mathcal{M}$) $< 0.01$
    return $\mathcal{M}$
end procedure
```

Algorithm 2 Rank-weighted neighbor updates

```
procedure ApplyRankWeightedUpdates($\mathbf{x}, \mathcal{M}, \mathcal{N}_k$)
    for $j = 1 \dots k$ in $\mathcal{N}_k$ (rank-ordered) do
        $w_j \leftarrow 2^{-(j-1)}$
        $\mathbf{node} \leftarrow \mathcal{N}_k[j]$
        $\mathbf{e} \leftarrow \mathbf{x} - \mathbf{node.w}$
        $\mathbf{node.m} \leftarrow (1 - \alpha w_j)\mathbf{node.m} + \alpha w_j\, \mathbf{e}$; $\mathbf{node.s} \leftarrow (1 - \alpha w_j)\mathbf{node.s} + \alpha w_j\, \mathbf{e}^{\circ 2}$
        $\rho \leftarrow \|\mathbf{node.m}\|/(\mathbf{node}.\sigma + \varepsilon)$; $\mathbf{node.a} \leftarrow \mathbf{node.a} + \eta_{cent}\, \rho\, \mathbf{node.m}$
        UpdateOja($\mathbf{node.u}, \mathbf{e}$)
        $\mathbf{node.h} \leftarrow \mathbf{node.h} + w_j$    ▷ Fractional hits and link counters
    end for
end procedure
```

Algorithm 3 Simplex-native updates (Stage 2)

```
procedure ApplySimplexNativeUpdates($\mathbf{x}, \mathcal{C}$)
    $C \leftarrow$ LocateWinningSimplex($\mathbf{x}$)    ▷ Barycentric or nearest with projection
    Order vertices of $C$ by increasing $\|\mathbf{x} - \mathbf{w}\_v\|$; assign rank-weights $w_j = 2^{-(j-1)}$
    for each vertex $v$ of $C$ in rank order do
        $\mathbf{e} \leftarrow \mathbf{x} - \mathbf{w}\_v$; $\mathbf{m}\_v \leftarrow (1 - \alpha w_j)\mathbf{m}\_v + \alpha w_j\, \mathbf{e}$; $\mathbf{s}\_v \leftarrow (1 - \alpha w_j)\mathbf{s}\_v + \alpha w_j\, \mathbf{e}^{\circ 2}$
        $\rho \leftarrow \|\mathbf{m}\_v\|/(\sigma\_v + \varepsilon)$; $\mathbf{a}\_v \leftarrow \mathbf{a}\_v + \eta\_cent\, \rho\, \mathbf{m}\_v$
    end for
    Accumulate facet pressures for $C$ per S5
end procedure
```

S12. Theoretical Guarantees

S12.1 Expressivity and Approximation

Let p be a probability density on a compact set $\Omega \subset \mathbb{R}^d$ with a (piecewise) Lipschitz continuous log-density and bounded curvature of its principal level sets. For any $\varepsilon>0$, there exist finite budgets for recursion depth, mesh resolution, and non-linear warps (mini-NSFs) such that the model density q learned by the Proteus procedure satisfies $\|p - q\|_{L_1} < \varepsilon$.

The proof is constructive, following the algorithmic stages. We show that the total approximation error can be bounded by the sum of errors from a hierarchical decomposition, a piecewise-linear approximation of each component, and a non-linear residual correction. By making the resolution and capacity of each stage sufficient, the total error can be made arbitrarily small.

S12.2 Correspondence: Simplex Equilibrium Implies Effective Node Equilibrium

Setting and assumptions. Consider a simplicial complex with nondegenerate simplex stars around each node, locally stationary sampling, and a fixed characteristic scale τ^*. Let $P(V_i)$ denote Voronoi mass at node i and $P(C)$ the mass of simplex C.

Claim. Driving the system toward simplex equilibrium ($P(C) \approx 1/M$ for all simplexes) induces effective node equilibrium ($P(V_i) \approx 1/N$ up to junction effects and sampling noise).

Sketch. Suppose a node i has excess mass $P(V_i) \gg 1/N$. Samples concentrate near w_i, biasing residuals in the adjacent simplex star $\mathcal{S}(i)$. The barycentric residual means in $\mathcal{S}(i)$ acquire outward components, raising torsional stress and face pressures that cannot be canceled by PL translation alone. The Torsion Ladder therefore either (i) performs a torsion-aligned split creating a child j in the outward direction, or (ii) fits a local mini-NSF that unbends the region and reduces the asymmetric flux. In both cases the effective domain of V_i contracts in favor of V_j or of neighboring stars, strictly decreasing $P(V_i)$ while increasing mass in adjacent cells. Repeating this argument shows that any persistent Voronoi over-mass produces simplex-level stresses whose resolution decreases the imbalance, establishing a negative feedback toward node equilibrium. Deviations persist only at structural junctions (e.g., different intrinsic dimensions), where systematic bimodality and asymmetric links form an informative signature rather than a failure.

Remarks. The argument uses only local stationarity and nondegenerate stars; concentration bounds convert it to a finite-sample statement. Empirically, the correspondence manifests around overloaded nodes as: sharply bimodal simplex activity within the node's star at dimensionality junctions; a clean partition of link statistics into high-significance backbone links versus low-significance bridge links; and bridge links that exhibit persistent directional hit asymmetry together with low net face pressure, distinguishing stable junctions from active frontiers.

The Signature of a Dimensionality Junction. The "informative mismatch" at structural junctions manifests as a concrete, stable, and observable statistical signature. A node located where, for example, a 1D filament meets a 2D sheet will persistently exhibit:

- **Bimodal Simplex Activity:** The distribution of internal activity values for its adjacent simplexes will be sharply bimodal.
- **Divergent Link Significance:** Its links will partition into high-significance "backbone" links (within a manifold) and low-significance "bridge" links (connecting manifolds).
- **Stable Asymmetry on Bridge Links:** Bridge links show high, stable directional hit-counter asymmetry but, crucially, low net face pressure. This distinguishes a stable junction from an active, growing frontier, where both asymmetry and pressure would be high.

This emergent signature is not a failure of the algorithm, but a feature that allows the system to report on the complex local topology of the data.

S13. Bayesian Non-Parametric View of the Proteus Pipeline

S13.1 Objects and Geometry-Induced Conditionals

Let $X = \{x_n\}_{n=1}^N \subset \mathbb{R}^d$ and let $G = (\mathcal{V}, \mathcal{E})$ denote the Hebbian/GNG graph with node locations $\{z_i\}$. Per-node BMU hit counts are n_i and directional edge counts are $n_{i\to j}$ (first/second NN "conditional hits"). Define empirical conditionals $\hat{p}(j \mid i) = n_{i\to j}/\max(1, n_i)$.

Let $\mathcal{S}$ denote the dual (Delaunay-like) complex with simplexes $S \in \mathcal{S}$ that tile the space between adjacent nodes. We introduce unknown simplex masses $m = \{m_S\}_{S\in\mathcal{S}}$ with $m_S \geq 0$ and $\sum_S m_S = 1$. Geometry induces barycentric weights $w_{iS} \in [0,1]$ with $\sum_{S\ni i} w_{iS} = 1$, and the mass "seen" by node i is

$$M_i(m) = \sum_{S\ni i} w_{iS}\, m_S\,.$$

A geometry-derived conditional from $i \to j$ implied by m (one sensible choice) is

$$q(j \mid i; m) = \frac{\sum_{S\ni i,j}\ \kappa_{ijS}\, m_S}{\sum_{S\ni i}\ \kappa_{iS}\, m_S}\,,$$

with positive geometric terms κ reflecting shared face length/area/solid angle or conductance.

S13.2 Likelihood from Dirichlet–Multinomial Evidence

Treat outgoing counts at node i as multinomial draws given $q(\cdot \mid i; m)$:

$$\{n_{i\to j}\}_j \,\big|\, n_i,\, m \sim \mathrm{Multinomial}\big(n_i,\, q(\cdot \mid i; m)\big)\,.$$

Placing a small Dirichlet smoothing prior on $q(\cdot \mid i)$ and integrating it out yields a Dirichlet–multinomial factor at node i,

$$\phi_i(m) \ \propto \prod_{j\in\mathcal{N}(i)} q(j \mid i; m)^{\,n_{i\to j}}\,,$$

up to a gamma-ratio constant independent of m. Thus each node contributes a log-linear potential in $\log q(j \mid i; m)$: Hebbian counts provide conjugate evidence about local transitions, while m is the global mass field that must rationalize all local transitions via geometry.

S13.3 Priors on the Mass Field m

Several Bayesian priors are natural, each imparting a different character to the model:

- **Simplex-wide Dirichlet (composition).** $m \sim \mathrm{Dirichlet}(\beta_S)_{S\in\mathcal{S}}$. Captures uncertainty on a fixed partition. It is conjugate if counts were observed directly in cells; here, the link is indirect via the transition likelihood, but the model remains coherent.
- **Logistic-Gaussian MRF (smoothness).** Let $\theta_S \in \mathbb{R}$ with GMRF prior $\theta \sim \mathcal{N}(0,\, \tau^{-1}L^{+})$ on the dual-graph Laplacian L, and set $m_S = \exp(\theta_S)/\sum_{S'} \exp(\theta_{S'})$. This is a logistic-normal prior that encourages neighboring simplexes to have similar mass, inducing spatial smoothness.
- **Dirichlet/Pólya tree on a mesh (hierarchical composition).** Impose a hierarchical split scheme over $\mathcal{S}$ (e.g., via a spanning tree) and place Dirichlet priors at each split. This defines a non-parametric model over recursive partitions, where the partition itself is learned by the Hebbian process.
- **Normalized random measures.** Assign independent $\tilde{m}_S \sim \mathrm{Gamma}(a_S, 1)$ and set $m_S = \tilde{m}_S/\sum_{S'} \tilde{m}_{S'}$. This is equivalent to a Dirichlet with parameters a_S on a fixed set $\mathcal{S}$ and connects the model to the broader theory of completely random measures.

S13.4 Posterior as a Factor Graph and Sum–Product

The posterior over m (up to normalization) is

$$p(m \mid \mathrm{data}) \;\propto\; \underbrace{\prod_{i\in\mathcal{V}} \phi_i(m)}_{\text{Dirichlet–multinomial edge evidence}} \;\times\; \underbrace{\psi(m)}_{\text{prior on } m},$$

where ψ is one of the priors above. Sum–product (belief propagation) on the dual factor graph performs approximate marginalization of this posterior. With a logistic-GMRF prior, this becomes a CRF-like model with spatial smoothness; with a Dirichlet prior, it is a compositional model without coupling.

S13.5 MaxEnt/MAP Connection

Setting a flat prior on m and maximizing the log-likelihood yields the MaxEnt/MLE estimator

$$\hat{m} \;=\; \arg\max_{m\in\Delta} \sum_i \sum_j n_{i\to j}\, \log q(j \mid i; m)\,.$$

Adding $\log \psi(m)$ yields a MAP estimator under the chosen Bayesian prior. Thus the Proteus MaxEnt + sum–product pipeline is the zero-prior (or uniform-prior) MAP limit of this coherent Bayesian model.

S13.6 Relation to BNP Families and Multiscale Refinement

The Proteus framework synthesizes ideas from several BNP families, but its data-driven, geometric construction creates key differences.

At a fixed scale, Proteus acts as a partition-based BNP model on a learned simplicial mesh. Its closest cousins are:

- **Pólya/Optional Pólya Trees (PT/OPT):** These models define densities on recursive, axis-aligned partitions. Proteus is similar in using a partition, but its partition is a *data-learned simplicial mesh with cycles*, not an axis-aligned tree. Furthermore, its likelihood is based on *transition counts* between nodes, not direct bin counts.

- **Log-Gaussian Cox Processes (LGCP):** These models use a latent Gaussian field to define a smooth log-density. Proteus with a logistic-GMRF prior is similar, but its likelihood is a CRF-like potential based on transition counts, coupling neighboring cells via $q(j \mid i; m)$, whereas LGCPs typically use direct Poisson counts in fixed bins.

- **Dependent Dirichlet Processes (DDP):** These models allow mass allocation to vary over space. However, DDPs typically randomize mixture weights and atoms in an infinite-dimensional space, whereas Proteus defines a *finite-dimensional* mass vector m whose components are geometrically coupled by the learned Hebbian topology.

Across scales, the coarse→fine recursion resembles a hierarchical BNP prior. For a coarse cell $P \in \mathcal{S}^{(s)}$ with children $\mathrm{ch}(P) \subset \mathcal{S}^{(s+1)}$, the refinement can be modeled as a stochastic process:

$$m_{\mathrm{ch}(P)}^{(s+1)} \mid m_P^{(s)} \sim \mathrm{Dir}\big(\alpha_s \, \pi_{\mathrm{ch}(P)}^{(s)} \, m_P^{(s)}\big), \qquad \sum_{C \in \mathrm{ch}(P)} m_C^{(s+1)} = m_P^{(s)} .$$

This is the multiscale Dirichlet refinement used by Pólya trees, but applied to a simplicial mesh. The warm-start procedure in Proteus, where the converged state at scale s initializes the search at $s+1$, is a deterministic analogue of using the posterior at one level as the prior for the next. The variance-based splitting rule acts as a data-driven control on mesh refinement, analogous to a Bayes factor test for splitting a cell.

S13.7 Practical Notes

Consistency and identifiability. As data grow, $n_{i \to j}$ stabilize and the posterior over m concentrates, provided the geometric router $q(\cdot \mid i; m)$ is sufficiently injective in the local masses (choose κ accordingly). A weak smoothness prior (e.g., logistic-GMRF) can also ensure identifiability. **Uncertainty propagation.** A fully Bayesian treatment propagates finite-sample noise in $n_{i \to j}$ into credible intervals over the mass field m. The MaxEnt/MAP approach provides a point estimate, which can be seen as the zero-prior ($\tau \to 0$) or infinite-data limit.

S13.8 Summary of Mapping

The Proteus pipeline can be mapped directly to a Bayesian non-parametric model:

- **Hebbian Learning/GNG** → A method for learning a **data-driven prior structure** (the simplicial partition $\mathcal{S}$).

- **Edge Hit Statistics** ($n_{i \to j}$) → Evidence for **Dirichlet–multinomial local likelihoods** ($\phi_i(m)$).

- **Dual-Graph Sum–Product** → An algorithm for **approximate Bayesian marginalization** over the global simplex mass field m.

- **MaxEnt Objective** → The **MAP inference** limit of the same model under a uniform prior.

This correspondence positions Proteus as a computationally efficient, geometrically grounded realization of BNP principles, using a deterministic, sequential MAP-tracking approach across scales in place of more expensive stochastic inference.

S14. Conditions for Equivalence of Stage 1 to Gaussian-Weighted PCA/SVD

Claim (informal). Under standard limits—frozen topology, infinitesimal step-size, stable neighborhoods, and infinite data—the Stage 1 moment tracker with rank-weighted updates and Oja's rule performs *local* Gaussian-weighted PCA/SVD at node locations. The variance threshold acts as a scale-dependent cap that adaptively refines regions whose Gaussian-windowed variance exceeds the target.

S14.1 Setting and Kernel Mapping

Let K_σ denote an isotropic Gaussian window with bandwidth σ, and let w_i be a (temporarily fixed) node location. Using the scale calibration in S2.1, set

$$\sigma \;=\; \sqrt{\tau}/c_{d,k}\,,$$

so that the per-node statistics at w_i target the Gaussian-smoothed local moments at the Stage 1 threshold τ.

Define the Gaussian-windowed local mean and covariance centered at w_i:

$$\mu_i \;=\; \mathbb{E}_{K_\sigma(\cdot - w_i)}[x]\,, \qquad C_i \;=\; \mathbb{E}_{K_\sigma(\cdot - w_i)}\big[(x-\mu_i)(x-\mu_i)^\top\big].$$

S14.2 Conditions for Equivalence

The following constraints yield the reduction of Stage 1 statistics to Gaussian-weighted PCA/SVD:

1. **Frozen topology and slow geometry.** During the estimation window, disable splits/pruning and take "slow" geometric motion (deferred moves with $\eta_{\text{cent}} \to 0$) so positions change on a slower time scale than moment convergence (cf. S2.3 deferred nudge).

2. **Infinitesimal step-size, infinite data.** Use $\alpha \to 0$ with a Robbins–Monro schedule or a small constant EWMA weight (S2.2) and an effectively infinite stream of i.i.d. samples from $p(x)$. Then EWMAs converge to the corresponding kernel expectations.

3. **Stable, approximately isotropic neighborhoods.** Use ANN settings where k is large enough and efSearch sufficiently high so the rank-ordered neighbor sets approximate samples from an isotropic ball around w_i of radius $\mathcal{O}(\sigma)$. The rank-geometric weights $2^{-(j-1)}$ then act as a discrete radial kernel that weakly converges to K_σ as $N \to \infty$, $k \to \infty$, $k/N \to 0$.

4. **Oja's rule for the principal subspace.** Apply standard Oja updates to residuals $e = x - w_i$ (cf. S8.1). Under the above sampling and step-size limits, Oja converges almost surely to the principal subspace of C_i; the tracked eigenpairs match those from the SVD of the local Gaussian-weighted covariance.

5. **Variance-cap thresholding.** Use the split criterion $\sigma_i^2 > \tau_{\text{local},i}$ with $\tau_{\text{local},i}$ mapped from the control per S2.4. With EWMAs equal to kernel moments, this is equivalent to adaptively refining where $\operatorname{tr}(C_i)$ exceeds the target scale.

S14.3 Consequences

Under these conditions, the Stage 1 per-node estimates equal the Gaussian-windowed moments at w_i:

$$m_i \to \mu_i - w_i\,, \qquad \sigma_i^2 \to \mathrm{tr}(C_i)\,; \quad \text{Oja}(e) \;\Rightarrow\; \text{eigvectors/values of } C_i.$$

Thus, the procedure performs *local PCA/SVD at scale σ*, and the thresholded growth rule implements an adaptive partition until the local Gaussian-smoothed variance falls below the cap.

S14.4 Explicit Limits Recovering SVD

- **Global PCA limit.** Single node, $k \to N$, and σ larger than the data diameter ⇒ uniform weights over the dataset; the covariance reduces to the global covariance and Oja recovers global PCA.
- **Local PCA limit.** Fixed nodes, $\alpha \to 0$, large k with isotropic neighborhoods ⇒ each node estimates the SVD of C_i, the Gaussian-windowed local covariance at w_i.
- **Piecewise-linear manifold limit.** With growth enabled at fixed τ, the mesh partitions the space so each cell maintains $\mathrm{tr}(C_i) \le \tau_{\mathrm{local},i}$; within cells, local SVD is recovered and provides the dominant direction for geometric refinement (via Oja).

S14.5 What Is *Not* the Same

- With active growth/topology changes, the method is *not* a single global SVD of the convolved dataset; it is a stitched collection of local Gaussian-weighted SVDs with adaptive centers and supports.
- At finite k, rank-geometric weights only approximate Gaussian radial weighting; exact equality holds asymptotically under the neighborhood stability conditions.
- Stage 2 constructs (simplicial complex, torsion audit, dual flow) have no SVD reduction; the equivalence pertains solely to Stage 1 at fixed scale.

S14.6 Takeaway

With the calibration $\sigma = \sqrt{\tau}/c_{d,k}$ (S2.1) and the constraints above, Stage 1 reduces to **local Gaussian-weighted PCA/SVD**, while the variance threshold implements **scale-aware adaptive refinement**. This provides a precise bridge between the Hebbian/GNG dynamics and kernel PCA at a chosen characteristic scale.