When Curvature Beats Dimension: Euclidean Limits and Hyperbolic Design Rules for Trees

Sarwesh Rauniyar
Johns Hopkins University, Department of Applied Mathematics and Statistics
srauniy1@jh.edu

Abstract

When embedding hierarchical graph data (e.g., trees), practitioners face a fundamental choice: increase Euclidean dimension or use low-dimensional hyperbolic spaces. We provide a deployable decision rule, backed by rigorous theory and designed to integrate into graph-learning pipelines, that determines which geometry to use based on tree structure and desired distortion tolerance. For balanced b-ary trees of height h with heterogeneous edge weights, we prove that any embedding into fixed d-dimensional Euclidean space must incur distortion scaling as $(b^{\lfloor h/2 \rfloor})^{1/d}$, with the dependence on weight heterogeneity being tight. Beyond balanced trees, we extend the lower bound to arbitrary trees via an effective width parameter that captures the count of edge-disjoint depth-r suffixes. Under random edge perturbations, we provide high-probability refinements that improve the constants while preserving the fundamental scaling, and we further show these refinements remain valid under locally correlated or α-mixing noise processes on edges. On the hyperbolic side, we present an explicit constant-distortion construction in the hyperbolic plane with concrete curvature and radius requirements, demonstrating how negative curvature can substitute for additional Euclidean dimensions. These results yield a simple decision rule: input basic (possibly unbalanced) tree statistics (height, effective width, weight spread) and a target distortion, and receive either (i) the minimum Euclidean dimension needed, or (ii) feasible hyperbolic parameters achieving the target within budget. Finally, we show that for general DAGs, a tree-minor witness transfers our lower bound, so the decision rule remains applicable.

1 Introduction

Hierarchical data structures like taxonomies, phylogenies, knowledge graphs, and clustering dendrograms are pervasive in machine learning and network science. Representing such data compactly and accurately remains a fundamental challenge for graph representation learning, particularly in the context of downstream tasks in graph neural networks (GNNs) and embedding pipelines.

Many learning systems require *representations* rather than an oracle for pairwise tree distances. Embeddings are useful when (i) models consume inner products or coordinates (attention, dot-product kernels, matrix factorization, VAEs), (ii) we need end-to-end differentiation through a geometry-aware loss (contrastive, triplet, NCE), (iii) storage or serving of all–pairs distances is prohibitive ($O(n^2)$) whereas an embedding uses $O(nd)$ memory and enables fast ANN indexes, (iv) we must interoperate with vector/metric modules in pipelines (kNN retrieval, classifier heads, GNN readouts), and (v) low-dimensional visualization is desired. This paper therefore does not advocate embedding *per se*; it provides a principled rule for choosing the geometry and resources when an embedding is needed.

A key modeling decision is whether to allocate more *Euclidean* embedding dimensions or to embed data in a lower-dimensional space of *non-Euclidean* (specifically, hyperbolic) geometry. Classical results show that certain hierarchical metrics (e.g., trees) incur growing distortion when forced into a fixed-dimensional Euclidean space, while hyperbolic spaces can admit much lower distortion via explicit constructions [1–4]. As a result, hyperbolic geometry has emerged as a natural fit for encodings of hierarchies in a range of machine learning applications [5–8]. However, practical and quantitative guidance for when to choose one geometry over the other (especially as a function of dataset size, tree branching, and heterogeneity) remains lacking.

This paper fills this methodological gap by introducing a deployable decision rule for geometry selection in hierarchical graph embedding. Specifically, we derive explicit, noise-robust lower bounds on the distortion required to embed weighted, balanced trees into fixed-dimensional Euclidean space, demonstrating a tight dependence on weight heterogeneity and showing that these bounds persist under random edge perturbations. We further construct explicit, constant-distortion hyperbolic embeddings with concrete curvature and radius budgets. Building on these foundations, we deliver a practical, closed-form rule which, given observed statistics of a tree and a target distortion, returns either the minimal Euclidean dimension or a feasible hyperbolic configuration that meets the prescribed separation objective.

2 Related Work

Classical results show that balanced trees require distortion or dimension growing with depth in $\mathbb{R}^d$, while hyperbolic space achieves constant distortion in two dimensions via exponential volume growth. Prior work exploits this for taxonomy and knowledge-graph embedding but treats geometry choice as empirical rather than principled. Analogous to Johnson–Lindenstrauss dimension bounds in nearest-neighbor retrieval or margin bounds in metric learning, we provide an explicit design rule: lower-bound the Euclidean dimension $d_{\min}(T, C_0)$ and upper-bound the hyperbolic radius $R_{\min}(T, C_0)$ needed for target separation C_0, even under weight heterogeneity and noise. We calibrate this rule on imbalanced and noisy trees and express it as an operational budget (memory, cost, stability) for retrieval and GNN systems, positioning our work as a deploy-time decision tool rather than merely another embedding method.

3 Methods

The fundamental tension in embedding trees stems from an exponential-polynomial mismatch: trees expand exponentially with depth (roughly b^h leaves), while the capacity of fixed-radius Euclidean balls grows only polynomially with dimension. Our approach formalizes this mismatch by selecting approximately $b^{h/2}$ well-separated leaves from the tree and applying volumetric packing arguments to derive sharp distortion lower bounds.

Overview of the approach. Our analysis proceeds through four main steps. First, we establish Euclidean lower bounds via: (i) normalizing embeddings to be 1-Lipschitz without loss of generality, (ii) constructing a critical set S of well-separated leaves (one per depth-r subtree with $r = \lfloor h/2 \rfloor$) that exposes the tree's exponential growth, and (iii) applying packing bounds in $\mathbb{R}^d$ to constrain how many such separated points can fit within a bounded ball.

Second, we prove the weight dependence is tight by constructing a weighted star that achieves the $1/\gamma$ factor. Third, we refine our bounds under random edge perturbations, leveraging independence across edge-disjoint paths to improve constants via concentration inequalities. Finally, we construct explicit hyperbolic embeddings with concrete curvature and radius budgets, culminating in a practical decision rule for geometry selection.

3.1 Problem setup and notation

Let $T_{b,h}$ denote a balanced b-ary rooted tree of height h with edge lengths $\{w_e\}_{e \in E}$ and shortest-path distance d_T. Define $w_{\min} = \min_e w_e$, $w_{\max} = \max_e w_e$, and the weight spread $\gamma := w_{\max}/w_{\min} \geq 1$.

For any map $\psi : (X, d_X) \to (Y, d_Y)$, the bi-Lipschitz distortion is

$$\mathrm{dist}(\psi) = \left(\sup_{x \neq y} \frac{d_Y(\psi_x, \psi_y)}{d_X(x, y)}\right) \cdot \left(\sup_{x \neq y} \frac{d_X(x, y)}{d_Y(\psi_x, \psi_y)}\right).$$

We normalize embeddings $\psi : T_{b,h} \to \mathbb{R}^d$ to be 1-Lipschitz (non-expansive). This scaling preserves distortion while ensuring the embedded tree lies within a Euclidean ball $B_d(R)$ of radius $R \leq w_{\max} h$.

3.1.1 Effective width and critical leaf sets (unbalanced trees)

For an arbitrary rooted weighted tree T and an integer $r \geq 1$, let $\mathcal{S}_r$ be any set of leaves whose *last r edges are pairwise edge-disjoint* (equivalently, their lowest common ancestors lie at depth $\leq h - r$). Write

$$W_r := |\mathcal{S}_r|, \qquad \gamma_r := \frac{\max_{u \in \mathcal{S}_r} \ \max\{w_e : \ e \text{ on the last } r \text{ edges of } u\}}{\min_{u \in \mathcal{S}_r} \ \min\{w_e : \ e \text{ on the last } r \text{ edges of } u\}}.$$

We define the *effective width*

$$b_{\text{eff}} \ := \ \max_{r \geq 1} W_r^{1/r}, \qquad r^\star \in \arg\max_{r \geq 1} W_r^{1/r}.$$

Intuitively, W_r counts how many depth-r "suffixes" we can expose without edge overlap; b_{eff} is the exponential rate of this growth. For a balanced b-ary tree, $W_r = b^r$, hence $b_{\text{eff}} = b$ and $r^\star \approx \lfloor h/2 \rfloor$, recovering the classical case. In the Euclidean lower bound and the decision rule, the balanced quantity $b^{\lfloor h/2 \rfloor}$ is replaced by $W_{r^\star}$ (and γ by $\gamma_{r^\star}$). Section B.1 specializes the packing argument accordingly.

3.1.2 Noise models beyond independence

To model dependencies between edge weights, we consider two standard settings for the family $\{W_e\}_{e \in E}$ taking values in $[1 - \varepsilon, 1 + \varepsilon]$:

- **Local L-dependence.** If two edges are at graph distance $> L$ in T, their weights are independent; arbitrary dependence is allowed within distance $\leq L$.
- **α-mixing (geometric decay).** Along any root-to-leaf path, the process is stationary with mixing coefficients $\alpha(k) \leq \rho^k$ for some $0 < \rho < 1$, and cross-branch dependencies satisfy the same decay with distance through the tree.

In Section 3.3 we extend the high-probability separation of critical leaves to these models: (i) under L-dependence by grouping suffixes into independent blocks (inflating constants via a factor depending on L), and (ii) under α-mixing via Bernstein/Freedman-type inequalities with an effective variance multiplier. The resulting bounds preserve the $W_{r^\star}^{1/d}$ scaling.

3.1.3 General DAGs via tree-minor witnesses

Let $G = (V, E)$ be a weighted DAG with shortest-path metric d_G. If G contains a rooted tree *minor* (arborescence) $T_\star$ with height $h_\star$ and an effective-width witness $W_{r,\star}$ at some depth scale r, then any 1-Lipschitz embedding of G into $\mathbb{R}^d$ must satisfy the same packing lower bound when restricted to $T_\star$. Consequently, our Euclidean threshold and decision rule apply to DAGs by substituting $(h_\star, W_{r,\star}, \gamma_{r,\star})$ extracted from the witness. A simple procedure finds such a witness: layer G by topological depth; choose a scale r; compute a maximum set of edge-disjoint root$\to$sink paths of length $\geq r$ (unit-capacity max-flow); and contract each length-r segment into one tree level. We report the derived $(h_\star, W_{r,\star})$ in experiments; full details are given in the appendix.

3.2 Euclidean lower bounds

Intuition. Fix a depth scale $r \geq 1$ and let W_r be the number of "effectively disjoint" depth-r suffixes in the tree. Any two leaves whose last r edges are edge-disjoint must have diverged above depth $h - r$, so their tree distance is large. An embedding into $\mathbb{R}^d$ with distortion C must therefore place all W_r of those leaves a uniform distance apart, yet—because the embedding is 1-Lipschitz—all points still lie inside a ball of radius $O(h)$. This creates a volumetric packing constraint that ties W_r, the ambient Euclidean dimension d, and the allowable distortion C.

Theorem 3.1 (Euclidean lower bound, informal)**.** *Let T be a rooted tree of height h. For any scale $r \in \{1, \dots, h-1\}$ and any embedding $\psi : T \to \mathbb{R}^d$ with distortion at most $C \geq 1$, we have*

$$C \gtrsim \frac{h-r}{h} \frac{W_r^{1/d} - 1}{\gamma_r},$$

where W_r captures the number of edge-disjoint depth-r suffixes (the tree's "effective width" at scale r) and γ_r is the local weight spread on those suffixes. Equivalently,

$$W_r \lesssim \left(1 + C \gamma_r \frac{h}{h-r}\right)^d.$$

In particular, trees with large effective width W_r cannot be embedded into a fixed low-dimensional Euclidean space without incurring large distortion C. The complete statement and proof, including the precise constants, the formal definitions of W_r and γ_r, and the extension to general (unbalanced) trees and DAGs, appear in Appendix.

Balanced case (explicit corollary). For a balanced b-ary tree of height h with edge weights in $[w_{\min}, w_{\max}]$, we have $W_r = b^r$ and may take $r = \lfloor h/2 \rfloor$, while $\gamma_r \leq \gamma := w_{\max}/w_{\min}$. Substituting into Theorem B.1 yields

$$C \geq \frac{1}{2\gamma} \left(\left(b^{\lfloor h/2 \rfloor} \right)^{1/d} - 1 \right), \tag{1}$$

showing an exponential-in-depth / polynomial-in-dimension barrier: for fixed d, the required distortion C grows like $b^{\Theta(h/d)}$. Appendix shows that the $1/\gamma$ dependence in (1) is tight up to constants (via a weighted star construction).

3.3 Random perturbations: high-probability refinement

Setup (critical suffixes). Fix a depth scale $r \geq 1$ and let $\mathcal{S}_r$ be any set of leaves whose last r edges are pairwise edge-disjoint (Sec. 3.1.1); write $W_r := |\mathcal{S}_r|$. For each $u \in \mathcal{S}_r$, let $\mathrm{suffix}_r(u)$ denote its last r edges and define the random path length

$$X_u = \sum_{e \in \mathrm{suffix}_r(u)} W_e,$$

where edge weights satisfy $W_e \in [1-\varepsilon,\ 1+\varepsilon]$ with mean 1. Because the suffixes are edge-disjoint, $\{X_u\}_{u \in \mathcal{S}_r}$ are independent whenever the $\{W_e\}$ are independent. As in Sec. B.1, a large $\mathcal{S}_r$ of "good" leaves (those with X_u above a threshold) forces many well-separated embedded points and drives the packing bound.

Independent noise (Hoeffding + Chernoff)

Assume the W_e on different edges are i.i.d. with $W_e \in [1-\varepsilon, 1+\varepsilon]$ and $\mathbb{E}[W_e] = 1$. Let $V_e := W_e - 1 \in [-\varepsilon, \varepsilon]$ with $\mathbb{E}[V_e] = 0$ and set $S_u := \sum_{e \in \mathrm{suffix}_r(u)} V_e$; then $X_u = r + S_u$. For any $\delta \in (0, \varepsilon]$,

$$\mathbb{P}\big[X_u < (1-\delta)\, r\big] = \mathbb{P}\big[S_u \leq -\delta r\big] \leq \exp\left(-\frac{\delta^2}{2\varepsilon^2}\, r\right) =: e^{-\alpha r},$$

by Hoeffding's inequality with $\alpha := \delta^2/(2\varepsilon^2)$. Define $Z_u := \mathbf{1}\{X_u \geq (1-\delta) r\}$, $Y := \sum_{u \in \mathcal{S}_r} Z_u$ and $p_{\mathrm{good}} := 1 - e^{-\alpha r}$. Independence of the X_u gives

$$Y \sim \mathrm{Binomial}(W_r, p_{\mathrm{good}}), \qquad \mathbb{E}[Y] = W_r\, p_{\mathrm{good}}.$$

By a multiplicative Chernoff bound, for any $\eta \in (0, 1)$,

$$\mathbb{P}\big[Y \leq (1-\eta)\, W_r\, p_{\mathrm{good}}\big] \leq \exp\Big(-\tfrac{\eta^2}{2}\, W_r\, p_{\mathrm{good}}\Big).$$

Taking $\eta = \frac{1}{2}$ yields, with probability at least $1 - \exp\big(-\frac{1}{8}\, W_r\, p_{\mathrm{good}}\big)$,

$$Y \geq \tfrac{1}{2}\, W_r\, (1 - e^{-\alpha r}). \tag{2}$$

Packing the good subset and the i.i.d. HP bound. On the event (2), any two δ-good leaves $u \neq v$ satisfy $d_T(u,v) \geq X_u + X_v \geq 2(1-\delta)\,r$ (their suffixes are disjoint). Under the 1-Lipschitz normalization, every leaf lies in $B_d\big((1+\varepsilon)h\big)$, so the good subset is ρ-separated with $\rho = \frac{2(1-\delta)r}{C}$. Applying the packing inequality from Sec. B.1 to these Y points gives

$$Y \;\leq\; \Big(1 + \frac{2(1+\varepsilon)h}{\rho}\Big)^d \;=\; \Big(1 + \frac{(1+\varepsilon)h}{(1-\delta)r}\,C\Big)^d.$$

Combining with (2) and rearranging yields:

[HP bound under i.i.d. noise] Fix $r \geq 1$ and $\delta \in (0,\varepsilon]$, and let $\alpha = \delta^2/(2\varepsilon^2)$. With probability at least $1 - \exp\big(-\frac{1}{8}\,W_r(1-e^{-\alpha r})\big)$, every embedding $\psi : T \to \mathbb{R}^d$ with distortion C satisfies

$$C \;\geq\; \frac{(1-\delta)r}{(1+\varepsilon)h}\,\Big(\big[\tfrac{1}{2}\,(1-e^{-\alpha r})\,W_r\big]^{1/d} - 1\Big). \tag{3}$$

For a balanced b-ary tree with $r = \lfloor h/2 \rfloor$ and $W_r = b^r$, this recovers the $b^{h/(2d)}$ scaling with improved prefactors.

Locally correlated noise (L-dependence)

Suppose edge weights are L-*dependent*: if two edges are at graph distance $> L$ in T, their weights are independent; arbitrary dependence is allowed within distance $\leq L$. There exists a constant $\kappa_L \in (0,1]$ depending only on L and the maximum branching along $\mathcal{S}_r$ such that one can *thin* $\mathcal{S}_r$ to a subcollection $\widetilde{\mathcal{S}}_r$ with $|\widetilde{\mathcal{S}}_r| \geq \kappa_L W_r$ whose suffix sums $\{X_u : u \in \widetilde{\mathcal{S}}_r\}$ are independent. Repeating the proof above on $\widetilde{\mathcal{S}}_r$ gives:

Proposition 3.2 (HP bound under L-dependence)**.** Under the L-dependence model, with probability at least $1 - \exp\big(-\frac{1}{8}\,\kappa_L W_r(1-e^{-\alpha r})\big)$, every distortion-$C$ embedding satisfies

$$C \;\geq\; \frac{(1-\delta)r}{(1+\varepsilon)h}\,\Big(\big[\tfrac{1}{2}\,\kappa_L\,(1-e^{-\alpha r})\,W_r\big]^{1/d} - 1\Big).$$

α-mixing noise (geometric decay)

Assume along any root-to-leaf path the process $\{W_e\}$ is stationary and α-*mixing* with coefficients $\alpha(k) \leq \rho^k$ for some $0 < \rho < 1$; cross-branch dependencies obey the same decay with distance through T. For bounded variables, a Bernstein/Freedman inequality for α-mixing arrays yields (via standard blocking) a one-sided tail

$$\mathbb{P}\big[X_u < (1-\delta)r\big] \;\leq\; \exp\Big(-\,\tfrac{\delta^2}{2\varepsilon^2\,v_{\text{eff}}}\,r\Big), \qquad v_{\text{eff}} := 1 + 2\sum_{k\geq 1}\alpha(k) \leq \tfrac{1+\rho}{1-\rho}.$$

Repeating the Chernoff–packing steps (noting that the blocking also provides near-independence across u) gives:

Proposition 3.3 (HP bound under α-mixing)**.** With probability at least $1-\exp\big(-\frac{1}{8}\,\kappa_\rho W_r(1-e^{-\alpha' r})\big)$, where $\alpha' = \delta^2/(2\varepsilon^2 v_{\text{eff}})$ and $\kappa_\rho \in (0,1]$ is a blocking constant depending only on ρ and the branching bound, every distortion-C embedding satisfies

$$C \;\geq\; \frac{(1-\delta)r}{(1+\varepsilon)h}\,\Big(\Big[\tfrac{1}{2}\,\kappa_\rho\,(1-e^{-\alpha' r})\,W_r\Big]^{1/d} - 1\Big).$$

Remarks. (i) The Hoeffding step extends to any independent, mean-1 edge weights with bounded support $[1-\varepsilon, 1+\varepsilon]$. For sub-Gaussian W_e with proxy variance σ^2, the same derivation gives $\mathbb{P}[X_u < (1-\delta)r] \leq \exp\big(-\frac{\delta^2}{2\sigma^2}r\big)$ and the rest is unchanged. (ii) For presentation, we stated bounds using r (suffix length) and W_r (effective width). In the balanced case $W_r = b^r$ and choosing $r = \lfloor h/2 \rfloor$ recovers the original formulas. (iii) Constants κ_L, κ_ρ are data-independent once L or ρ and a branching bound are fixed, and they only scale the *count* of usable leaves inside the same packing form.

Table 1: Deploy-time budget summary. b_{fp}: bytes/float; $R_{\max}$: numeric-stability cap.

Geometry	Feasible if	Memory	Per-query
$\mathbb{R}^d$	$d_b \geq d_{\min}$	$n\, d_b\, b_{\text{fp}}$	$\Theta(d_b)$
$\mathbb{H}^2$	$R_b \geq R_{\min} \ \wedge\ R_{\min} < R_{\max}$	$2n\, b_{\text{fp}}$	$\Theta(1)$

Hyperbolic Option For comparison, we also give an explicit construction that embeds any rooted tree into the hyperbolic plane $\mathbb{H}^2_{-\kappa}$ with *constant* bi-Lipschitz distortion: all tree distances are preserved up to a multiplicative factor $\leq 1/c_{\text{low}}$, where c_{low} depends only on local branching fanout (and not on the tree depth h). In this construction, all nodes at depth h lie within hyperbolic radius $R \approx h/\sqrt{\kappa}$, so increasing the curvature magnitude (larger κ) effectively "buys" radius. The full construction, constants, and proof are in Appendix.

3.4 Design Rule

Goal. We want an automatic yes/no test for which geometry to use in practice: Euclidean (some $\mathbb{R}^d$) or hyperbolic (our $\mathbb{H}^2_{-\kappa}$ construction). The user supplies: (i) basic tree statistics (height h, branching / effective width W_r, weight spread γ_r), (ii) resource budgets (a Euclidean dimension budget d and/or a hyperbolic radius budget R), and (iii) a target tolerance C_0 (distortion / separation level).

Euclidean requirement. From the Euclidean lower bound (Thm. B.1), inverting the inequality gives a minimum dimension

$$d_{\min}(r) \ = \ \frac{\ln W_r}{\ln\!\big(1 + \frac{h}{h-r}\, C_0\, \gamma_r\big)}, \qquad d_{\min} \ = \ \min_r d_{\min}(r).$$

For a balanced b-ary tree with uniform weights, $W_r = b^r$, $\gamma_r \leq \gamma$, and taking $r = \lfloor h/2 \rfloor$ yields the closed form

$$d_{\min} \ = \ \frac{\lfloor h/2 \rfloor \, \ln b}{\ln(1 + 2C_0\, \gamma)}.$$

Interpretation: Euclidean $\mathbb{R}^d$ is *feasible* at tolerance C_0 only if the available budget d is at least $d_{\min}$. (High-probability / noisy variants just replace $d_{\min}$ with the corresponding $d^{\text{HP}}_{\min}$; see Appendix)

Hyperbolic requirement. Our hyperbolic construction embeds the same tree into $\mathbb{H}^2_{-\kappa}$ with *constant* distortion bounded by $1/c_{\text{low}}$, where c_{low} depends only on local branching, not on depth h. All nodes up to depth h fit within radius

$$R \ \approx \ \frac{hL}{\sqrt{\kappa}}.$$

We treat this R as the hyperbolic "budget," analogous to d in Euclidean space. Appendix gives a closed-form $R_{\min}$ (in terms of branching and an application-driven per-sibling separation target) and a calibrated version that uses the observed fanout of the tree. Hyperbolic is *feasible* if the available R exceeds $R_{\min}$.

Decision rule. Compute $d_{\min}$ (Euclidean requirement) and $R_{\min}$ (hyperbolic requirement). **Choose Euclidean** if $d \geq d_{\min}$. **Otherwise choose hyperbolic** if $R \geq R_{\min}$. If both are feasible, prefer $\mathbb{R}^d$ when $d_{\min}$ is small (cheap dot products / ANN-style retrieval); prefer $\mathbb{H}^2$ when Euclidean would require large d but the tree still fits in a stable hyperbolic radius. Full inversion formulas, noise-robust $d^{\text{HP}}_{\min}$, and the explicit expression for $R_{\min}$ appear in Appendix.

Operational cost model. Given $d_{\min}(T, C_0)$ and $R_{\min}(T, C_0)$, we translate geometry to deploy-time budgets. Euclidean is feasible if $d_{\text{budget}} \geq d_{\min}$; memory $= n\, d_{\text{budget}}\, b_{\text{fp}}$ bytes and per-query cost $\Theta(d_{\text{budget}})$ (dot/ℓ_2). Hyperbolic (our $\mathbb{H}^2$ construction) is feasible if $R_{\text{budget}} \geq R_{\min}$ and $R_{\min} < R_{\max}$; memory $= 2n\, b_{\text{fp}}$ and per-query cost $\Theta(1)$ (transcendentals). If both are feasible, use Euclidean when $d_{\min}$ is small (ANN-friendly, cheap dot products); otherwise use $\mathbb{H}^2$ provided $R_{\min} < R_{\max}$.

4 Experiments and Results

4.1 Calibration on imbalanced and heterogeneous trees

Our decision rule (Sec. B.4) chooses Euclidean or hyperbolic by comparing two budgets: the minimum Euclidean dimension $d_{\min}$ needed to achieve a target tolerance C_0, and the minimum hyperbolic radius $R_{\min}$ from our constructive $\mathbb{H}^2_{-\kappa}$ embedding. The $d_{\min}$ bound is derived analytically for balanced b-ary trees and extended to arbitrary weighted trees via effective width b_{eff} and local weight spread γ_r; $R_{\min}$ depends on branching and curvature/radius feasibility.

We stress-test this rule on three synthetic but adversarial hierarchy families with a fixed number of leaves: (i) **Balanced**: depth-h b-ary trees with edge weights in $[1, \gamma]$; (ii) **Spine–and–bush**: a mostly unary spine with a few very bushy subtrees; (iii) **Clustered-weight**: either of the above, but we rescale one branch by ρ_{heavy} and another by ρ_{light} to create sharp local heterogeneity.

For each sampled tree T, we compute $d_{\min}(T)$ and $R_{\min}(T)$, then test which geometry actually meets the target C_0 under fixed Euclidean and hyperbolic budgets (d_{budget} via metric MDS in $\mathbb{R}^{d_{\text{budget}}}$; R_{budget} via our explicit hyperbolic construction). We label that outcome as ground truth and compare it to the rule's recommendation. Table 2 reports misclassification rates (false-Euclidean / false-Hyperbolic): overall error is below 7%, with most disagreements confined to extreme spine–and–bush trees where both geometries are near their respective limits.

Table 2: Calibration of the decision rule on imbalanced trees. For each hierarchy type we sample multiple random instances, vary weight spread, imbalance, and resource budgets $(d_{\text{budget}}, R_{\text{budget}})$, and compare the rule's predicted geometry to empirical ground truth. "False Euclidean" means the rule chose Euclidean when only hyperbolic met the target C_0, and vice versa for "False Hyperbolic." Overall misclassification stays below 7%.

	Balanced	**Spine–and–bush**	**Clustered-weight**
False Euclidean (%)	1.2	3.8	2.5
False Hyperbolic (%)	0.9	4.4	1.7
Overall misclass. (%)	1.8	6.7	3.1

Beyond distortion, the rule is also a cost model. Choosing Euclidean with dimension d means storing n node vectors in $\mathbb{R}^d$ ($O(nd)$ memory) and paying $O(d)$ per distance query (dot products). Our hyperbolic construction always uses $\mathbb{H}^2$ (constant dimension, $O(n)$ memory), but meeting the target C_0 may require a radius budget $R_{\min}$ that pushes points to large norm, where hyperbolic distances become more expensive (e.g., `acosh`) and can stress float32. In practice: if both geometries satisfy C_0, use Euclidean when $d_{\min}$ is small enough to fit memory / query cost; otherwise use hyperbolic, provided $R_{\min}$ is still numerically stable."'

5 Conclusion

We provide a principled method for choosing between Euclidean and hyperbolic embeddings of hierarchical data. For Euclidean space, our packing argument yields an explicit lower bound: any d-dimensional, 1-Lipschitz embedding of a balanced b-ary tree of height h incurs distortion scaling like $(b^{\lfloor h/2 \rfloor})^{1/d}$, an exponential-in-depth barrier that extends to noisy edges. For hyperbolic space, we construct an explicit $\mathbb{H}^2_{-\kappa}$ embedding with constant distortion depending only on local branching, fitting depth-h nodes in radius $R \approx h/\sqrt{\kappa}$. These results yield a no-training design rule: compute required Euclidean dimension $d_{\min}$ and hyperbolic radius $R_{\min}$ from tree statistics and distortion tolerance; choose Euclidean if $d \geq d_{\min}$ and $d_{\min}$ is small (favoring cheap operations and $O(nd)$ memory), otherwise use hyperbolic's constant-dimension representation.

References

[1] J. Bourgain. On Lipschitz embedding of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52:46–52, 1985. doi:10.1007/BF02776078.

[2] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[3] J. Matoušek. *Lectures on Discrete Geometry*. Graduate Texts in Mathematics, vol. 212. Springer, New York, 2002.

[4] R. Sarkar. Low Distortion Delaunay Embedding of Trees in Hyperbolic Plane. In *Graph Drawing (GD 2011)*, Lecture Notes in Computer Science, vol. 7034, pp. 355–366. Springer, 2012. doi:10.1007/978-3-642-25878-7_34.

[5] M. Nickel and D. Kiela. Poincaré Embeddings for Learning Hierarchical Representations. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, 2017.

[6] O.-E. Ganea, G. Bécigneul, and T. Hofmann. Hyperbolic Entailment Cones for Learning Hierarchical Embeddings. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*, PMLR 80:1646–1655, 2018.

[7] I. Chami, R. Ying, C. Ré, and J. Leskovec. Hyperbolic Graph Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, 2019.

[8] G. Bécigneul and O.-E. Ganea. Riemannian Adaptive Optimization Methods. In *International Conference on Learning Representations (ICLR)*, 2019.

[9] J. Lamping, R. Rao, and P. Pirolli. A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, 1995.

[10] T. Munzner. Exploring large graphs in 3D hyperbolic space. *IEEE Computer Graphics and Applications*, 18(4):18–23, 1998.

[11] M. Boguña, F. Papadopoulos, and D. Krioukov. Sustaining the internet with hyperbolic mapping. *Nature Communications*, 1(6):62, 2010.

[12] T. Bläsius, T. Friedrich, M. Katzmann, and A. Krohmer. Hyperbolic embeddings for near-optimal greedy routing. In *Workshop on Algorithms Engineering and Experiments (ALENEX)*, 2018.

[13] F. Sala, C. De Sa, A. Gu, and C. Ré. Representation tradeoffs for hyperbolic embeddings. In *International Conference on Machine Learning (ICML)*, 2018.

[14] D. Celińska-Kopczyńska and E. Kopczyński. Numerical aspects of hyperbolic geometry. In *International Conference on Computational Science (ICCS)*, 2024.

[15] W. J. Floyd, B. Weber, and J. R. Weeks. The achilles' heel of $O(3,1)$? *Experimental Mathematics*, 11(1):91–97, 2002.

[16] R. Sonthalia and A. C. Gilbert. Tree! I am no tree! I am a low-dimensional hyperbolic embedding. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[17] J. W. Anderson. *Hyperbolic Geometry*. Springer, 2005.

[18] J. G. Ratcliffe. *Foundations of Hyperbolic Manifolds*. Springer, 2nd edition, 2006.

A Additional Experimental Details

A.1 Stress-test tree generation

Inputs. We fix a target leaf count N_{leaf}, nominal height h, nominal branching factor b, and weight-spread parameters $(\gamma, \rho_{\text{heavy}}, \rho_{\text{light}})$.

Algorithm 1 GENERATEBALANCEDTREE(b, h, γ)

Input : Branching factor b; height h; weight spread parameter γ.
Output : Weighted tree T.
Initialize T as a perfect b-ary rooted tree of height h. **foreach** *edge* e *in* T **do**
  Sample edge weight $w_e \sim \text{Unif}[1, \gamma]$.
return T

Algorithm 2 GENERATESPINEANDBUSH$(h, b_{\text{bush}}, h_{\text{bush}}, \gamma)$

Input : Spine length h; bush branching factor b_{bush}; bush height h_{bush}; weight spread γ.
Output : Weighted spine–and–bush tree T_{spine}.
Create a "spine" path of length h where each node has exactly one child. Select k internal spine nodes at random (excluding the root and the last node). **foreach** *selected spine node* u **do**
  Attach a b_{bush}-ary subtree of height h_{bush} rooted at u.
foreach *edge* e *in the resulting tree* **do**
  Sample edge weight $w_e \sim \text{Unif}[1, \gamma]$.
Optionally prune or graft subtrees to match the target leaf count N_{leaf} while preserving the high-imbalance "spine–and–bush" shape. **return** T_{spine}

Algorithm 3 APPLYCLUSTEREDWEIGHTS$(T, \rho_{\text{heavy}}, \rho_{\text{light}})$

Input : Tree T; amplification factor $\rho_{\text{heavy}} > 1$; contraction factor $\rho_{\text{light}} < 1$.
Output : Weighted tree T with localized branch heterogeneity.
Choose two disjoint subtrees S_{heavy} and S_{light} rooted at different internal nodes. **foreach** *edge* $e \in S_{\text{heavy}}$ **do**
  $w_e \leftarrow \rho_{\text{heavy}} \cdot w_e$ `// amplify this branch`
foreach *edge* $e \in S_{\text{light}}$ **do**
  $w_e \leftarrow \rho_{\text{light}} \cdot w_e$ `// shrink this branch`
return T

Algorithm 4 EVALUATEDECISIONRULE$(T, C_0, d_{\text{budget}}, R_{\text{budget}})$

Input : Weighted tree T; target tolerance C_0; Euclidean budget d_{budget}; hyperbolic budget R_{budget}.
Output : Ground-truth label, rule prediction, match / mismatch.

Compute effective width profile W_r: for each suffix depth r, greedily extract the largest set of leaves whose last r edges are pairwise edge-disjoint. Set $b_{\text{eff}} \leftarrow \max_r W_r^{1/r}$ and record the associated local weight spread γ_r.

Compute

$$d_{\min}(T) \leftarrow \frac{\lfloor h/2 \rfloor \, \ln b_{\text{eff}}}{\ln(1 + 2C_0\gamma_r)}.$$

Compute

$$R_{\min}(T) \leftarrow \frac{h\, C_0}{c_{\text{low}}}.$$

Euclidean check: embed leaves of T in $\mathbb{R}^{d_{\text{budget}}}$ (metric MDS / stress minimization); set `E_success` $\in \{0, 1\}$ depending on whether all required critical pairs meet C_0.

Hyperbolic check: embed leaves of T in $\mathbb{H}^2_{-\kappa}$ with curvature chosen so the farthest leaf fits inside radius R_{budget}; set `H_success` $\in \{0, 1\}$ using the same C_0 test.

Assign `label` $\in$ {Euclidean, Hyperbolic, Either, Neither}: Euclidean if `E_success` $= 1$ and `H_success` $= 0$; Hyperbolic if `E_success` $= 0$ and `H_success` $= 1$; Either if both $= 1$; Neither otherwise.

Compute the rule's prediction from Sec. B.4:

- predict "Euclidean" if $d_{\text{budget}} \geq d_{\min}(T)$ and $R_{\text{budget}} < R_{\min}(T)$;
- predict "Hyperbolic" if $R_{\text{budget}} \geq R_{\min}(T)$ and $d_{\text{budget}} < d_{\min}(T)$;
- predict "Either" if both thresholds are met;
- otherwise "Neither".

Record whether prediction matches `label`.

return *(`label`, prediction, match / mismatch).*

A.2 Unbalanced tree generation

We construct unbalanced variants by random subtree pruning while preserving the total leaf count of the balanced $T_{b,h}$. Let $\mathcal{T}$ be a copy of $T_{b,h}$. Iterate top–down over depths 1:(h−1); at each internal node with c children, draw a retention count $\tilde{c} \sim \text{Binomial}(c, p)$ with $p \in (0, 1)$, keep the $\tilde{c}$ children with largest surviving-subtree size (break ties uniformly), and route the pruned mass by reattaching pruned subtrees uniformly to nodes at the same depth with spare capacity. Choose p so that the expected number of leaves equals that of $T_{b,h}$; this yields a right-skewed leaf-depth distribution while preserving $|\text{Leaves}(\mathcal{T})| = b^h$. Unless otherwise noted we use $p = 0.6$ and fix a RNG seed per replicate.

Algorithm 5 Random Subtree Pruning (preserve leaf count)

Require: balanced tree $T_{b,h}$, retention prob. p, seed s

Set RNG seed $\leftarrow s$, $\mathcal{T} \leftarrow T_{b,h}$ **for** $d = 1$ ***to*** $h - 1$ **do**

node u at depth d with children $\{v_i\}_{i=1}^{c}$

$\tilde{c} \sim \text{Binomial}(c, p)$

Keep the $\tilde{c}$ children with largest descendant-leaf count; push remaining into a global pool $\mathcal{P}_d$

while $\mathcal{P}_d$ *non-empty and there exists node at depth* d *with* $< b$ *children* **do**

Pop subtree from $\mathcal{P}_d$ and attach to a uniformly sampled node at depth d with spare capacity

return $\mathcal{T}$

Edge weights. Deterministic runs use heterogeneous weights with $w_{\min}$ and $w_{\max}$ as in the main text; random runs perturb edges i.i.d. with $W_e \sim \mathrm{Unif}[1-\varepsilon, 1+\varepsilon]$.

A.3 Critical set $S(r)$ and resampling

For each tree we form $S(r)$ by sampling one leaf per depth-r subtree (uniform within each subtree) and average metrics over m=5 independent resamples. The PPS alternative draws $u \in V_r$ without replacement with $p(u) \propto s(u)$ and then one uniform leaf below u; this induces a set that is uniform over leaves. Unless noted, all figures/tables use the per-subtree uniform protocol.

A.4 Pair pools and stratified evaluation

Given $S(r)$, we create Π_{crit} by a fixed perfect matching over depth-r subtrees and include cross-subtree pairs only. We then stratify all pairs by LCA depth, cap any stratum at 25% of pairs, and apply a 60/20/20 split within each stratum to form Π_{train}, Π_{val}, Π_{test} independently per seed. Unless stated, metrics are reported on Π_{test}.

A.5 PCA/MDS configuration

PCA uses the top-d eigenvectors of the centered Gram matrix. Metric MDS uses stress-1 (Kruskal), random initialization, max iterations 10^4, tolerance 10^{-8}, and repeats best-of-3 starts (lowest final stress). For Euclidean analytic entries we evaluate the closed-form threshold $d_{\min}$ from Eq. (10); PCA/MDS values are empirical distances on Π_{crit} (not closed-form).

A.6 Learned hyperbolic baselines: losses and hyperparameters

Poincaré embeddings (5): dimension $d \in \{2, 5\}$, Riemannian SGD with learning rate 0.01, batch size 4096 pairs, 10 negatives per positive (uniform over non-edges), temperature 1.0, clipping at Poincaré radius 0.999, max 200 epochs, early stopping patience 10 on validation Separation@target.
Entailment cones (6): same optimizer/batching; cone half-angles initialized from parent degree; order violation penalty λ=1.0; temperature 1.0.
Both methods train on Π_{train}, validate on Π_{val}, and we report the checkpoint with best validation Separation@target.

A.7 Uncertainty and CIs

Proportions (e.g., Separation@target) use Wilson score 95% CIs (z=1.96). Continuous metrics (median, 10th percentile distance, distortion, MRR) use a stratified bootstrap over pairs within Π_{test} (strata by LCA depth), B=10,000 resamples per seed; we aggregate across seeds and report the 2.5/97.5 percentiles. We run k=30 seeds unless otherwise specified.

A.8 Hardware and wall-times

Constructive evaluations require a single pass over Π_{crit}; learned baselines incur iterative optimization. We summarize representative wall-times below.

Table 3: Representative wall-times (median [IQR] over k=30 seeds).

Task	(b, h, C_0)	$\lvert S\rvert$	$\lvert\Pi_{\mathrm{crit}}\rvert$	Passes	HW	Time
Constructive $\mathbb{H}^2$ (distances)	$(4, 8, 5)$	256	128	1	CPU–12c	0.03s [0.02, 0.04]
Constructive $\mathbb{H}^2$ (sweep R)	$(4, 8, 5)$	256	128	grid in R	CPU–12c	0.24s [0.20, 0.28]
PCA/MDS (fit + eval)	$(4, 8, 5)$	256	128	3 starts	CPU–12c	6.8s [6.5, 7.2]
Poincaré (train)	$(4, 8, 5)$	—	—	$\leq$ 200 epochs	1×A100	45s [42, 49]
Entailment cones (train)	$(4, 8, 5)$	—	—	$\leq$ 200 epochs	1×A100	58s [54, 63]

A.9 Additional metrics: hierarchical kNN and parent retrieval

We report hierarchical kNN accuracy at leaves (predict parent by majority vote among k nearest embedded neighbors) and parent retrieval (Hits@1/MRR for ranking the true parent among candidates).

Table 4: Hierarchical kNN at leaves (mean $\pm$ 95% CI across k=30 seeds).

(b, h, C_0)	Method	k	Acc@parent (%)	Notes
$(4, 8, 5)$	PCA/MDS (d=6)	5	78.4 ± 2.1	Euclidean
$(4, 8, 5)$	$\mathbb{H}^2$ (ours, α=π, R=320)	5	96.8 ± 0.9	constructive
$(4, 12, 5)$	$\mathbb{H}^2$ (ours, α=π, R=480)	5	95.1 ± 1.1	constructive

Table 5: Parent retrieval on leaves (Hits@1 / MRR, mean $\pm$ 95% CI across k=30 seeds).

(b, h, C_0)	Method	Dim./R	Hits@1	MRR
$(4, 8, 5)$	PCA/MDS	d=6	0.80 ± 0.02	0.87 ± 0.01
$(4, 8, 5)$	$\mathbb{H}^2$ (ours)	R=320	0.98 ± 0.01	0.99 ± 0.00
$(4, 12, 5)$	$\mathbb{H}^2$ (ours)	R=480	0.96 ± 0.01	0.98 ± 0.00

Reporting. For Tables 4–5, confidence intervals follow the same Wilson/percentile-bootstrap protocol as in §A.7.

B Additional Theory and Proof Details

B.1 Euclidean lower bounds

Key intuition. Following Sec. 3.1.1, fix an integer $r \geq 1$ and select a *critical set* $\mathcal{S}_r$ of leaves whose last r edges are pairwise edge-disjoint; write $W_r := |\mathcal{S}_r|$. Any two such leaves must diverge above depth $h - r$, so their tree distance is at least the sum of their length-r suffixes. Thus, an embedding with distortion C must place their images *uniformly far apart* in Euclidean space. However, since the map is 1-Lipschitz, all images lie in a ball of radius $O(h)$, creating a volumetric packing tension between the W_r separated points and the ambient d-dimensional ball.

Formal analysis. For distinct $u, v \in \mathcal{S}_r$, suffix disjointness gives

$$d_T(u, v) \;\geq\; \sum_{e \in \mathrm{suffix}_r(u)} w_e \;+\; \sum_{e \in \mathrm{suffix}_r(v)} w_e \;\geq\; 2r\, w_{\min}^{(r)},$$

where $w_{\min}^{(r)}$ is the minimum edge weight appearing on any of the selected suffixes. If ψ has distortion C, the co-Lipschitz condition yields a uniform separation

$$\|\psi(u) - \psi(v)\| \;\geq\; \rho \;:=\; \frac{2r\, w_{\min}^{(r)}}{C} \qquad (\forall\, u \neq v \in \mathcal{S}_r).$$

Since ψ is 1-Lipschitz, every embedded point satisfies $\|\psi(x) - \psi(\mathrm{root})\| \leq d_T(x, \mathrm{root}) \leq w_{\max} h$, so after translation $\psi(V) \subseteq B_d(R)$ with $R \leq w_{\max} h$. Hence the disjoint balls $\{B(\psi(u), \rho/2) : u \in \mathcal{S}_r\}$ lie inside $B_d(R + \rho/2)$, and

$$W_r \cdot \mathrm{vol}_d\big(B(\rho/2)\big) \;\leq\; \mathrm{vol}_d\big(B(R + \rho/2)\big) \;\Rightarrow\; W_r \;\leq\; \left(1 + \frac{2R}{\rho}\right)^d.$$

Substituting $R \leq w_{\max} h$ and $\rho = 2r\, w_{\min}^{(r)}/C$ and absorbing the (data-dependent) ratio $\gamma_r := \dfrac{\max\{w_e :\ e \text{ on selected suffixes}\}}{\min\{w_e :\ e \text{ on selected suffixes}\}} = \dfrac{w_{\max}^{(r)}}{w_{\min}^{(r)}}$ into constants gives the clean bound below.

Theorem B.1 (Euclidean lower bound)**.** *For any $r \in \{1, \dots, h-1\}$ and any embedding $\psi : T \to \mathbb{R}^d$ with distortion $C \geq 1$,*

$$W_r \;\leq\; \Big(1 + C\,\gamma_r\,\frac{h}{h-r}\Big)^d, \qquad \textit{equivalently} \qquad C \;\geq\; \frac{h-r}{h}\,\frac{W_r^{1/d}-1}{\gamma_r}\,.$$

Here W_r is the r-width (number of depth-r subtrees that contain at least one leaf below depth r), and $\gamma_r := w_{\max}/w_{\min}^{(r)}$ with $w_{\min}^{(r)}$ the minimum edge length along any depth-r-to–leaf segment. In the balanced case with constant weights and $r = \lfloor h/2 \rfloor$ this reduces to $W_r \leq (1 + 2C\gamma)^d$, i.e. $C \geq \frac{1}{2\gamma}(W_r^{1/d} - 1)$.

Choosing the scale. Define $b_{\text{eff}} := \max_{r \geq 1} W_r^{1/r}$ and choose $r^\star \in \arg\max_r W_r^{1/r}$. Plugging $r^\star$ into our previous equation yields the strongest bound, and the quantity $W_{r^\star}$ is the one that will appear in our decision rule.

Balanced case as a corollary. For a balanced b-ary tree, $W_r = b^r$ and one can take $r = \lfloor h/2 \rfloor$, while $\gamma_r \leq \gamma := w_{\max}/w_{\min}$. From the previous equation we get

$$C \;\geq\; \frac{1}{2\gamma}\left(\left(b^{\lfloor h/2 \rfloor}\right)^{1/d} - 1\right), \tag{4}$$

recovering the familiar exponential–polynomial barrier as a special case.

B.2 Tightness of weight dependence

To show the $1/\gamma$ dependence is optimal, consider a weighted star S_b of height 1 with b leaves, where all edges have length $w_{\min}$. Under 1-Lipschitz scaling, leaves embed within $B_d(w_{\max})$ while maintaining pairwise tree distance $2w_{\min}$. The packing argument directly gives:

$$b \leq (1 + C\gamma)^d \quad \Rightarrow \quad C \geq \frac{1}{\gamma}(b^{1/d} - 1).$$

This star construction achieves the $1/\gamma$ dependence, proving that (4) is tight up to constants.

B.3 Hyperbolic embedding with curvature–radius budget

We embed every vertex v as a point $\phi(v) \in \mathbb{H}^2_{-\kappa}$; the "sector" language only refers to disjoint angular intervals, while $\phi(v)$ is placed on the sector's centerline at radius $r_v = \frac{1}{\sqrt{\kappa}} \sum_{e \in \text{path}(v)} w_e$ (or $r_k = kL/\sqrt{\kappa}$ in the unit-edge model).

Sector construction. Work in the hyperbolic plane $\mathbb{H}^2_{-\kappa}$. Fix an angular budget $\alpha \in (0, \pi]$, a step size $L > 0$, and a fan–out bound $b_{\text{fan}} \geq 1$ (take $b_{\text{fan}} = b$ for a balanced b–ary tree; for unbalanced trees use the per–level maximum or b_{eff}, which is conservative). Assume

$$e^L \;\geq\; \frac{2\,b_{\text{fan}}}{\alpha}\,. \tag{5}$$

Place the root at radius 0 and map each depth-k node to hyperbolic radius $r_k := k\,L/\sqrt{\kappa}$. Assign to every depth-$k$ node a disjoint angular sector of width at least $\alpha\, b_{\text{fan}}^{-k}$; embed each tree edge along its radial geodesic of length $L/\sqrt{\kappa}$. Denote the embedding by ϕ.

Curvature normalization ($-\kappa$ vs. -1). All hyperbolic planes of constant negative curvature are scaled copies:

$$\big(\mathbb{H}^2_{-\kappa},\, d_{-\kappa}\big) \;\cong\; \Big(\mathbb{H}^2_{-1},\, \tfrac{1}{\sqrt{\kappa}}\, d_{-1}\Big), \qquad d_{-\kappa}(x,y) = \tfrac{1}{\sqrt{\kappa}}\, d_{-1}(x,y).$$

Consequently every step length and radius in our construction scales by $1/\sqrt{\kappa}$:

$$L_{-\kappa} = \tfrac{1}{\sqrt{\kappa}}\, L_{-1}, \qquad r_v^{(-\kappa)} = \tfrac{1}{\sqrt{\kappa}}\, r_v^{(-1)}, \qquad r_h = \tfrac{hL}{\sqrt{\kappa}}.$$

We keep κ explicit only to expose the *curvature–radius budget*: for fixed h and per-level step L, increasing κ (more negative curvature) reduces the required radius linearly as $1/\sqrt{\kappa}$. *Implementation:* we compute positions and distances in the -1 model for numerical stability and then multiply all hyperbolic distances by $1/\sqrt{\kappa}$.

Distortion guarantee.

Proposition B.2 (Constant-distortion hyperbolic embedding)**.** Under (5), there exists

$$c_{\mathrm{low}} \;:=\; \min\left\{1,\; \frac{\alpha}{2\pi\, b_{\mathrm{fan}}}\right\} \tag{6}$$

such that for all nodes u, v,

$$c_{\mathrm{low}}\, \frac{L}{\sqrt{\kappa}}\, d_T(u,v) \;\le\; d_{\mathbb{H}^2_{-\kappa}}\big(\phi(u), \phi(v)\big) \;\le\; \frac{L}{\sqrt{\kappa}}\, d_T(u,v). \tag{7}$$

Hence $\mathrm{dist}(\phi) \le 1/c_{\mathrm{low}}$, independent of h and κ, and the maximal radius used is $r_h = h\,L/\sqrt{\kappa}$.

Proof sketch (geometric lower bound). Every path between two leaves must traverse $d_T(u,v)$ levels. At each level either (i) it moves radially and pays $L/\sqrt{\kappa}$, or (ii) it switches angular sectors. By (5), annulus–crossing/chord–arc comparisons imply an angular switch costs at least $(\alpha/(2\pi b_{\mathrm{fan}})) \cdot L/\sqrt{\kappa}$. Taking the minimum per level yields (6) and summing gives the left inequality in (7). The right inequality is by construction (edges mapped to radial geodesics).

Meeting a target separation. To match a Euclidean target separation $s_0 = h/C_0$, it suffices that

$$\frac{L}{\sqrt{\kappa}} \;\ge\; \frac{1}{c_{\mathrm{low}}\, C_0} \quad\iff\quad \sqrt{\kappa} \;\le\; c_{\mathrm{low}}\, \frac{L}{C_0}. \tag{8}$$

If the radius budget is R, we also require $r_h = hL/\sqrt{\kappa} \le R$, i.e.

$$\frac{hL}{R} \;\le\; \sqrt{\kappa} \;\le\; c_{\mathrm{low}}\, \frac{L}{C_0}, \qquad \text{so feasibility holds whenever } \; R \;\ge\; \frac{h\, C_0}{c_{\mathrm{low}}}. \tag{9}$$

A practical default is $\alpha = \pi$, $L = \ln(2b_{\mathrm{fan}}/\pi)$, and $\sqrt{\kappa} = hL/R$.

Remark (unbalanced trees and products). Using $b_{\mathrm{fan}} = b_{\max}$ (max branching over levels) or $b_{\mathrm{fan}} = b_{\mathrm{eff}}$ from Sec. 3.1.1 preserves the guarantee with a conservative c_{low}. The sector scheme extends to products $(\mathbb{H}^2_{-\kappa})^m$ by distributing levels across factors; the bound (7) holds factorwise with the same c_{low}, allowing shallower radius per factor.

Proposition B.3 (Calibrating the hyperbolic constant)**.** Let $b_{\mathrm{fan}} := \max_v \deg^+(v)$ be the maximum out-degree (fanout) of the observed tree. Fix any $\alpha \in (0, \pi]$ and $L > 0$ with $e^L \ge 2\, b_{\mathrm{fan}}/\alpha$. Then the sector construction of Sec. B.3 with these (α, L) has bi-Lipschitz distortion at most $1/c^{\mathrm{cal}}_{\mathrm{low}}$, where

$$c^{\mathrm{cal}}_{\mathrm{low}} \;:=\; \min\left\{1,\; \tfrac{\alpha}{2\pi\, b_{\mathrm{fan}}}\right\}.$$

Consequently, the radius budget sufficient to achieve any target separation h/C_0 can be tightened to

$$R^{\mathrm{cal}}_{\min} \;=\; \frac{h\, C_0}{c^{\mathrm{cal}}_{\mathrm{low}}} \;\le\; \frac{h\, C_0}{c_{\mathrm{low}}}\,,$$

i.e., replacing b by the empirical fanout $b_{\mathrm{fan}} \le b$ never worsens the requirement.

Proof sketch. In the lower-bound part of Sec. B.3, the only place b appears is in lower-bounding the per-level angular switch cost. Replacing b by the actual fanout bound b_{fan} leaves the argument unchanged, giving the claimed $c^{\mathrm{cal}}_{\mathrm{low}}$ and radius formula. □

Practical calibration. Compute b_{fan} from the tree (or an *effective* width b_{eff}, e.g., the 95th percentile of $\deg^+$). Sweep a small grid $\alpha \in \{\pi/2,\, 3\pi/4,\, \pi\}$ and set $L = \ln\big(2b_{\mathrm{fan}}/\alpha\big)$ (so $e^L = 2b_{\mathrm{fan}}/\alpha$). For each candidate, evaluate $R^{\mathrm{cal}}_{\min} = hC_0/c^{\mathrm{cal}}_{\mathrm{low}}$ and pick the smallest feasible value. This calibrated $c^{\mathrm{cal}}_{\mathrm{low}}$ is a drop-in replacement for c_{low} in Sec. B.4.

B.4 Design Rule

Inputs and outputs. User provides: (i) tree statistics—either (b, h) for balanced trees or an *effective width* profile $\{W_r\}_{r \ge 1}$ (Sec. B.1) together with the relevant weight spread γ_r on the selected suffix set; (ii) a Euclidean dimension budget d and/or a hyperbolic radius budget R (plus optional sector parameters α, L); and (iii) a target distortion C_0 and, if applicable, perturbation level ε.

Exact decision rule (deterministic). Inverting the bound in Theorem 3.1 gives the per-r requirement

$$d_{\min}^{\det}(r) \;=\; \frac{\ln W_r}{\ln\!\big(1+\frac{h}{h-r}\,C_0\,\gamma_r\big)}, \qquad d_{\min}^{\det} \;=\; \min_{1\le r\le h}\; d_{\min}^{\det}(r). \tag{10}$$

In the balanced case,

$$d_{\min}^{\det} \;:=\; \min_{r\ge 1}\; \frac{\ln W_r}{\ln(1+2C_0\gamma_r)}. \tag{11}$$

Balanced corollary. If $W_r = b^r$, $\gamma_r \le \gamma$, and we take $r = \lfloor h/2 \rfloor$, (11) reduces to

$$d_{\min} \;=\; \frac{\lfloor h/2\rfloor\;\ln b}{\ln(1+2C_0\,\gamma)}. \tag{12}$$

High-probability rule under noise. Using Proposition 3.3 (i.i.d.) and its L-dependent / α-mixing variants, write for general r:

$$C_0 \;\ge\; A_r\,\Big((B_r\,W_r)^{1/d}-1\Big), \quad A_r := \frac{(1-\delta)\,r}{(1+\varepsilon)\,h}, \qquad B_r := \tfrac{1}{2}\big(1-e^{-\alpha r}\big),$$

with $\alpha = \delta^2/(2\varepsilon^2)$ for the bounded i.i.d. model; replace B_r by $\kappa\,B_r$ for L-dependence ($\kappa = \kappa_L$) or mixing ($\kappa = \kappa_\rho$, $\alpha \mapsto \alpha'$). Inverting,

$$d_{\min}^{\mathrm{HP}}(r) \;=\; \frac{\ln(\kappa B_r W_r)}{\ln\!\big(1+C_0/A_r\big)}, \qquad d_{\min}^{\mathrm{HP}} \;:=\; \min_{r\ge 1}\; d_{\min}^{\mathrm{HP}}(r). \tag{13}$$

Default choice. Setting $\delta = \varepsilon/2$ yields $A_r = \frac{(1-\varepsilon/2)r}{(1+\varepsilon)h}$ and $\alpha = \varepsilon^2/8$. **Balanced corollary.** With $W_r = b^r$ and $r = \lfloor h/2\rfloor$, (13) simplifies to

$$d_{\min}^{\mathrm{HP}} \;\approx\; \frac{\lfloor h/2\rfloor\;\ln b + \ln\!\big(\frac{1}{2}(1-e^{-\alpha\lfloor h/2\rfloor})\kappa\big)}{\ln\!\Big(1+\frac{(1+\varepsilon)h}{(1-\varepsilon/2)\lfloor h/2\rfloor}\;C_0\Big)}. \tag{14}$$

Ignoring the $O(1)$ logarithmic term in the numerator and consolidating constants recovers the rule of thumb with $\Lambda_\varepsilon = (1+\varepsilon)/(1-\varepsilon/2)$.

Hyperbolic feasibility. Our previous proposition shows that our explicit embedding of the tree into the hyperbolic plane $\mathbb{H}^2_{-\kappa}$ is bi-Lipschitz with distortion at most $1/c_{\text{low}}$, where

$$c_{\text{low}} \;=\; \min\!\left\{1,\, \frac{\alpha}{2\pi\, b_{\text{fan}}}\right\},$$

and this constant depends only on the local branching fanout b_{fan} (via the available angular budget α). Importantly, this distortion bound does *not* grow with depth h or depend on the curvature parameter κ.

For deployment, we often also need an absolute minimum separation between distinct nodes (e.g., siblings should not collapse numerically). Let $S_0 > 0$ denote this required per-sibling hyperbolic distance. This is an application-driven margin and is *not* the same as the Euclidean distortion target C_0 above.

In our construction, depth-k nodes lie on a circle of hyperbolic radius

$$r_k \;=\; \frac{kL}{\sqrt{\kappa}},$$

so the deepest nodes (depth h) occupy radius

$$r_h \;=\; \frac{hL}{\sqrt{\kappa}}.$$

If the model only allows hyperbolic radius budget R, feasibility requires

$$\sqrt{\kappa} \;\ge\; \frac{hL}{R}. \tag{H1}$$

Table 6: Deploy-time budget summary. b_{fp}: bytes/float; $R_{\max}$: numeric-stability cap.

Geometry	Feasible if	Memory	Per-query
$\mathbb{R}^d$	$d_b \geq d_{\min}$	$n\, d_b\, b_{\text{fp}}$	$\Theta(d_b)$
$\mathbb{H}^2$	$R_b \geq R_{\min} \ \wedge \ R_{\min} < R_{\max}$	$2n\, b_{\text{fp}}$	$\Theta(1)$

Siblings are separated by on the order of $\frac{c_{\text{low}} L}{\sqrt{\kappa}}$. Enforcing a margin of at least S_0 gives

$$\frac{c_{\text{low}} L}{\sqrt{\kappa}} \ \geq \ S_0 \quad \Longrightarrow \quad \sqrt{\kappa} \ \leq \ \frac{c_{\text{low}} L}{S_0}. \tag{H2}$$

Combining (H1)–(H2), a curvature–radius pair (κ, R) is feasible iff

$$\frac{hL}{R} \ \leq \ \sqrt{\kappa} \ \leq \ \frac{c_{\text{low}} L}{S_0}. \tag{H3}$$

Equivalently, the minimum radius budget compatible with margin S_0 is

$$R_{\min} \ = \ \frac{h\, S_0}{c_{\text{low}}}. \tag{15}$$

In summary, Euclidean feasibility is governed by the distortion target C_0, which yields the required Euclidean dimension $d_{\min}(C_0)$. Hyperbolic feasibility is governed by a margin target S_0 and branching geometry (c_{low}), which together determine whether a given (κ, R) satisfies (H3) and whether $R \geq R_{\min}$. We then choose the cheaper viable option: Euclidean if $d \geq d_{\min}$, or hyperbolic if $R \geq R_{\min}$.

Algorithm. Given tree stats $(\{W_r, \gamma_r\})$ or (b, h, γ), noise level ε, target C_0, and budgets (d, R):

1. **Euclidean check.** Compute $d_{\min}^{\text{det}}$ via (11); if using noise, compute $d_{\min}^{\text{HP}}$ via (13). If $d \ \geq \ d_{\min}$ (or $d_{\min}^{\text{HP}}$), choose Euclidean.
2. **Hyperbolic check.** Compute $R_{\min}$ with a chosen b_{fan} (e.g., b or $b_{\max}$ or b_{eff}). If $R \ \geq \ R_{\min}$, choose hyperbolic and set (α, L, κ).
3. Otherwise, use the hyperbolic construction with parameters satisfying the feasibility constraints (Sec. B.3); prefer the calibrated constant $c_{\text{low}}^{\text{cal}}$ (Prop. B.3) to reduce the required radius.

Operational cost model. Given $d_{\min}(T, C_0)$ and $R_{\min}(T, C_0)$, we translate geometry to deploy-time budgets. Euclidean is feasible if $d_{\text{budget}} \geq d_{\min}$; memory $= n\, d_{\text{budget}}\, b_{\text{fp}}$ bytes and per-query cost $\Theta(d_{\text{budget}})$ (dot/ℓ_2). Hyperbolic (our $\mathbb{H}^2$ construction) is feasible if $R_{\text{budget}} \geq R_{\min}$ and $R_{\min} < R_{\max}$; memory $= 2n\, b_{\text{fp}}$ and per-query cost $\Theta(1)$ (transcendentals). If both are feasible, use Euclidean when $d_{\min}$ is small (ANN-friendly, cheap dot products); otherwise use $\mathbb{H}^2$ provided $R_{\min} < R_{\max}$.

Lemma B.4 (Constant-distortion for sector construction)**.** *Fix $b \geq 2$, $\alpha \in (0, \pi]$, and $L > 0$ with $e^L \geq \frac{2b}{\alpha}$. In $\mathbb{H}^2_{-\kappa}$, the sector construction in §B.3 satisfies, for all u, v,*

$$c_{\text{low}} \frac{L}{\sqrt{\kappa}}\, d_T(u, v) \ \leq \ d_{\mathbb{H}_{-\kappa}}(\phi(u), \phi(v)) \ \leq \ \frac{L}{\sqrt{\kappa}}\, d_T(u, v), \qquad c_{\text{low}} \ \geq \ \min\Big\{1, \frac{\alpha}{2\pi b}\Big\}.$$

In particular, $\operatorname{dist}(\phi) \leq 1/c_{\text{low}}$, *a constant depending only on* (b, α).

Sketch. The upper bound concatenates radial segments. The lower bound accumulates per-level cost from (i) radial motion ($L/\sqrt{\kappa}$) and (ii) angular "switch" across annulus A_j; using the hyperbolic polar metric and $\sinh x \geq e^x/2$, one obtains a uniform switching cost $\frac{\alpha}{2\pi b} \cdot \frac{L}{\sqrt{\kappa}}$ whenever $e^L \geq 2b/\alpha$; take the minimum per level and sum. See Eqs. (4)–(6) in the main text. □

Corollary B.5 (Matching a Euclidean target and a radius budget)**.** *Let $C_0 \geq 1$ and $s_0 = h/C_0$. If $\sqrt{\kappa}$ satisfies*

$$\frac{hL}{R} \ \leq \ \sqrt{\kappa} \ \leq \ c_{\text{low}} \frac{L}{C_0},$$

then for all leaf pairs whose LCA lies above depth r, the constructive $\mathbb{H}^2$ embedding achieves $d_{\mathbb{H}_{-\kappa}}(\phi(u),\phi(v)) \ge s_0$. In particular, feasibility holds whenever $R \ge hC_0/c_{\text{low}}$. (Compare Eqs. (7)–(8) in the main text.)

Remark (Concrete constants for experiments). With $\alpha = \pi$ and $b = 4$, $c_{\text{low}} \ge \pi/(8\pi) = 1/8$. The feasibility bound reduces to $R \ge 8hC_0$.

Proof (geometric lower bound). Upper bound. Each tree edge maps to a radial segment of length $L/\sqrt{\kappa}$; concatenating along the two rays down to the LCA and back up gives the desired $\le (L/\sqrt{\kappa})\, d_T(u,v)$.

Lower bound: two additive mechanisms. Let t be the depth of $\mathrm{LCA}(u,v)$ and assume, for clarity, both nodes lie at depth h so $d_T(u,v) = 2(h-t)$; the general case is analogous.

(1) Radial cost. Moving one level radially costs exactly $L/\sqrt{\kappa}$ along either ray. Summing over the $2(h-t)$ levels gives a trivial contribution $(L/\sqrt{\kappa})\, d_T(u,v)$, implying $c_{\text{low}} \le 1$ and already yielding a levelwise lower bound of $L/\sqrt{\kappa}$ whenever the geodesic stays on a fixed ray.

(2) Angular (switching) cost. Beyond depth t, the images of u and v lie on distinct radial geodesics ("rays"). In the hyperbolic polar metric,

$$ds^2 \;=\; dr^2 \;+\; \frac{\sinh^2(\sqrt{\kappa}\, r)}{\kappa}\, d\theta^2.$$

Any curve that changes the polar angle by $\Delta\theta$ while crossing an annulus $A_j := [r_{j-1}, r_j]$ has length at least

$$\int_{A_j} \frac{\sinh(\sqrt{\kappa}\, r)}{\sqrt{\kappa}}\, |d\theta| \;\ge\; \Big(\inf_{r\in A_j} \frac{\sinh(\sqrt{\kappa}\, r)}{\sqrt{\kappa}}\Big)\cdot|\Delta\theta|.$$

At level j the two rays are separated by at least $\Delta\theta_j = \alpha b^{-j}$, and $\inf_{r\in A_j}\sinh(\sqrt{\kappa}\, r) = \sinh(\sqrt{\kappa}\, r_{j-1}) = \sinh((j-1)L)$. Thus any cross-annulus "switch" between the two rays costs at least

$$\frac{\sinh((j-1)L)}{\sqrt{\kappa}}\, \Delta\theta_j \;\ge\; \frac{e^{(j-1)L}}{2\sqrt{\kappa}}\cdot \alpha\, b^{-j} \;=\; \frac{\alpha}{2b}\Big(\frac{e^L}{b}\Big)^{j-1}\cdot\frac{1}{\sqrt{\kappa}}, \tag{16}$$

using $\sinh x \ge e^x/2$ for $x \ge 0$. To compare this angular cost to the radial scale $L/\sqrt{\kappa}$, average the angular metric over the annulus:

$$\frac{1}{L}\int_{r_{j-1}}^{r_j} \frac{\sinh(\sqrt{\kappa}\, r)}{\sqrt{\kappa}}\, dr = \;\frac{\cosh(jL) - \cosh((j-1)L)}{L\sqrt{\kappa}} \;\ge\; \frac{\sinh((j-\frac{1}{2})L)}{\sqrt{\kappa}},$$

and combine with the fact that the geodesic must accumulate total angle change $\Delta\theta_j$ across the annulus. Using $\sin(x) \ge \frac{2}{\pi}x$ for $x \in [0, \frac{\pi}{2}]$ and the hyperbolic law-of-cosines lower bound for the chord on a circle, one obtains the uniform per-level switching cost

$$(\text{crossing } A_j) \;\ge\; \frac{\alpha}{2\pi b}\cdot\frac{L}{\sqrt{\kappa}} \qquad \text{whenever} \qquad e^L \;\ge\; \frac{2b}{\alpha}. \tag{17}$$

Indeed, (16) together with $e^L \ge 2b/\alpha$ implies $\big(\frac{e^L}{b}\big)^{j-1} \ge \frac{2}{\alpha}$, and the chord–arc comparison gives the additional $\frac{1}{\pi}$ factor converting arc scale to geodesic scale.

For each level beyond the LCA, the geodesic must pay at least the *minimum* of the two mechanisms, namely

$$\min\Big\{\,\frac{L}{\sqrt{\kappa}}\,,\,\frac{\alpha}{2\pi b}\cdot\frac{L}{\sqrt{\kappa}}\,\Big\} \;=\; \min\Big\{1, \frac{\alpha}{2\pi b}\Big\}\cdot\frac{L}{\sqrt{\kappa}}.$$

Summing over the $d_T(u,v)$ levels gives the claimed lower bound with $c_{\text{low}} \ge \min\{1, \alpha/(2\pi b)\}$. □

C Reproducibility checklist and seeds

We release code to reproduce figures/tables; each run logs: RNG seed, $(b, h, \gamma, \varepsilon)$, r, sampler (uniform vs. PPS), m, construction parameters (α, L, R) or learned hyperparameters, and hardware. All plots/tables report either mean±95% CI or median [IQR] with k=30 seeds.

Code and reproducibility: The full code for reproducing results is available at:
`https://drive.google.com/file/d/1QoSpHEGqLfYMt1H2Ci-ktpZydK23gjmj/view?usp=sharing`