On Hamming–Lipschitz Type Stability of the Subdominant (Minmax) Ultrametric: Theory and Simple Proofs

Anonymous authors
Paper under double-blind review

Abstract

We study the subdominant (minmax) ultrametric as an operator on pairwise data. Prior stability results show that this operator is non-expansive under uniform perturbations in the supremum norm and in the Gromov–Hausdorff sense, but they say nothing about how widely sparse, targeted edits can ripple through the hierarchy. We close this gap with a pair-count Lipschitz theory in Hamming space: we bound how many ultrametric entries can change, regardless of their magnitudes. The analysis is routed through the *minimum spanning tree* (MST), which encodes the ultrametric as path bottlenecks. Our first theorem proves a locality principle; only pairs whose MST path crosses an edited or newly exposed cut can change, so the impact is confined to a union of fundamental cut rectangles. Building on this, we derive an instance dependent ℓ_0 type Lipschitz bound whose constant is determined entirely by the MST's exposed cuts. We then show optimality by constructing cases where a single off-tree edit forces a quadratic number of changes, so no smaller universal constant is possible for our proposed Lipschitz constant. Finally, under a mild minimal-overlap condition, the upper bound on the number of changed entries of the ultrametric is order-tight, yielding a two-sided characterization of propagation. Conceptually, this advances a magnitude-versus-extent picture for ultrametric stability: classical results control how much entries move under uniform perturbation; our theory controls how far changes spread under sparse edits. Additionally, as a proof of concept, we derive a risk score from our Lipschitz constant that identifies vulnerable edges in the graph. We use this score to drive two case studies: vulnerability maps of deep embeddings of CIFAR-10, ImageNet-10, and STL-10, where targeted edits to high-score edges cause far larger ultrametric and clustering changes than random edits with the same budget, and fragility maps in a superpixel-based single image segmentation that highlight load-bearing boundaries.

1 Introduction

Hierarchical clustering (Ward Jr, 1963) provides a fundamental way to represent relational data through nested partitions and dendrograms (Shepard, 1962). Among all possible hierarchies, the subdominant (or minmax) ultrametric (Sibson, 1971; Hartigan, 1985; Jain & Dubes, 1988) occupies a canonical position: it is the subdominant ultrametric that minimally dominates the given dissimilarity and serves as the unique projection from the space of pairwise distances onto the cone of ultrametrics. Formally, given a metric space (X, d), the minmax ultrametric can be viewed as the output of an operator $U : (X, d) \mapsto (X, u_d)$, where $u_d : X \times X \mapsto \mathbb{R}$ is the minmax ultrametric associated with d. This operator, central to both metric geometry and modern representation learning, transforms arbitrary dissimilarities into tree-structured metrics that preserve the strongest pairwise connections. Carlsson et al. (2010) established that the dendrogram of *single linkage clustering* (SLC) and the minmax ultrametric are equivalent representations of the same hierarchical structure. This quantity coincides exactly with the merge height of the two points in the single-linkage dendrogram. Hence, single-linkage hierarchical clustering can be viewed as computing the *maximal ultrametric dominated by the original distances*, providing a precise geometric correspondence between dendrograms and ultrametrics. Further, Carlsson et al. (2010) formulated a rigorous mathematical framework for hierarchical clustering by identifying it with a mapping from finite metric spaces to ultramet-

ric spaces. Their key theoretical result is that the minmax ultrametric map is 1-Lipschitz (non-expansive) with respect to Gromov-Hausdorff ($\mathcal{GH}$) metric. Thus, the minmax ultrametric (dendrogram of SLC) is non-expansive under arbitrary perturbations of the input metric in $\mathcal{GH}$ sense, implying that small changes in distances cannot amplify in the induced ultrametric. In contrast, other linkage-based operators such as complete or average linkage, fail to satisfy this non-expansive property. Their stability theorem establishes single linkage as the unique hierarchically consistent and Lipschitz stable ultrametric projection. Regarding assumptions on perturbations, their analysis is fully general; no probabilistic or noise model is imposed. The only requirement is that the metric perturbation be bounded in the Gromov–Hausdorff sense, meaning that all pairwise distances between the two metric spaces differ by at most a small additive ε. Under this assumption, every ultrametric distance changes by at most ε. Thus, their stability theorem captures uniform, global perturbations of the metric, but does not address sparse or localized adversarial edits.

Chowdhury et al. (2016) further analyzed stability in more concrete norms. They prove that the minmax ultrametric operator $U : (X, d) \mapsto (X, u_d)$ is 1-Lipschitz under the sup norm. Mathematically, $\| u_d - u_{\tilde{d}} \|_\infty \leqslant \| d - \tilde{d} \|_\infty$, for all metrics $d, \tilde{d}$ on X. Crucially, they formalized a duality between Gromov's tree embedding and the ultrametric structure produced by SLC. They introduced a measure of deviation from ultrametricity, quantifying how far a finite metric space is from being perfectly treelike. Through this duality, they proved that the single-linkage dendrogram computes the ultrametric that minimizes additive distortion up to a bound depending on the space's ultrametricity and doubling dimension. In essence, the single-linkage dendrogram corresponds to an optimal ultrametric tree embedding whose distortion reflects both the local *ultrametricity* of the data and its intrinsic dimensional complexity. Together, these results position the single-linkage dendrogram as both an optimal low-distortion tree embedding and a globally stable (Lipschitz-continuous) map from metric data to hierarchical structure.

Recently, Mikhailov (2025) extended the stability result of Carlsson et al. (2010) to the full Gromov–Hausdorff class of all (possibly unbounded) metric spaces. He showed that the canonical subdominant (min–max) ultrametricization mapping $U : (X, d) \mapsto (X, u_d)$, obtained from the Carlsson–Memoli construction, is *1-Lipschitz* with respect to the Gromov–Hausdorff distance, not only on bounded spaces but on arbitrary metric spaces. This viewpoint interprets U as a non-expansive map between *clouds*, i.e., equivalence classes of spaces at finite Gromov–Hausdorff distance. Moreover, for any dotted connected metric space A, he exhibited an inverse relationship between U and Cartesian products with A: on the cloud of bounded ultrametric spaces, the map $\Psi : X \mapsto X \times A$ is an isometric embedding and $d_{\mathcal{GH}}\big(U(\Psi(X)), X\big) = 0$, so that $U(\Psi(X))$ is (Gromov–Hausdorff) isometric to X, while ultrametric spaces are precisely the fixed points of U. Conceptually, this places the ultrametricization operator as a globally Lipschitz-stable transformation over the Gromov–Hausdorff landscape, further linking metric geometry with hierarchical clustering.

1.1 Gaps in Previous Theory and Motivation

Existing stability results for the minimax ultrametric u_d address *uniform* perturbations, proving non-expansiveness in the entrywise ℓ_∞ metric, and via standard comparisons, in the Gromov–Hausdorff framework. These guarantees bound the *magnitude* of change everywhere but are agnostic to the *sparsity* and *locality* of the perturbation: a single large edit renders $\|d - \tilde{d}\|_\infty$ large, and the resulting bound allows every entry of u_d to move by that amount, even when the true effect is confined to a tiny portion of pairs. In particular, the uniform theory provides no control over the *extent* (support size) of the induced change in u_d when perturbations are sparse.

We close this gap by analyzing the ultrametric map $U : (X, d) \mapsto (X, u_d)$ in a Hamming-type setting on pairwise dissimilarity matrices. Concretely, for a finite metric d on X, we encode d as its upper-triangular distance vector and equip this space with the Hamming metric

$$d_H(d, \tilde{d}) \;=\; \#\big\{\{x, y\} \subseteq X : d(x, y) \neq \tilde{d}(x, y)\big\}, \tag{1}$$

i.e., the number of pairs whose dissimilarities are edited. Equivalently, $d_H(d, \tilde{d}) = \|d - \tilde{d}\|_0$, where $\| \cdot \|_0$ counts nonzero coordinates. Although $\| \cdot \|_0$ is not a norm, it induces the bonafide Hamming metric d_H, and we establish a sparsity-sensitive Lipschitz-type bound $\|u_d - u_{\tilde{d}}\|_0 \;\leqslant\; L_T^\star \, \|d - \tilde{d}\|_0$, which limits *how widely* changes can propagate through the ultrametric (the support of the deviation), independently of their

amplitudes. $L_T^\star$ depends only on pairs whose MST path crosses an exposed cut; in the worst case on an n-node graph, $L_T^\star \leqslant \binom{n}{2}$. Thus, our result complements the classical $\ell_\infty/\mathcal{GH}$ non-expansiveness along an orthogonal axis: the prior theory controls *how much* entries may move under uniform noise, while our Hamming-metric guarantee controls *how many* entries can change under sparse edits. Together, these yield a magnitude-versus-extent stability picture that was previously unavailable for the ultrametric operator.

Continuity (Lipschitz) Type	**Noise Model / Assumptions**	**Lipschitz expression**	**Reference**
ℓ_∞-continuity (sup-norm)	Bounded entrywise perturbations of all pairs.	$\|u_d - u_{\tilde{d}}\|_\infty \leqslant \|d - \tilde{d}\|_\infty$	Carlsson et al. (2010); Dey et al. (2017)
$\mathcal{GH}$ stability	Arbitrary perturbation measured in $\mathcal{GH}$-distance.	$d_{\mathcal{GH}}(u_d, u_{\tilde{d}}) \leqslant d_{\mathcal{GH}}(d, \tilde{d})$	Carlsson et al. (2010)
$\mathcal{GH}$ (semi-)stability	Stable in $\mathcal{GH}$ only when the input metric is (nearly) ultrametric.	$d_{\mathcal{GH}}(u_d, u_{\tilde{d}}) \leqslant d_{\mathcal{GH}}(d, \tilde{d})$	Martínez-Pérez (2015)
ℓ_∞ (restricted perturbation)	Additive perturbation on a subset S'(or insertion/removal of points).	$\|u_d - u_d^{S'}\|_\infty \leqslant \|d - d^{S'}\|_\infty$	Chowdhury et al. (2016)
$\mathcal{GH}$ stability for unbounded spaces	Entrywise-bounded perturbations for not-necessarily bounded metric spaces.	$d_{\mathcal{GH}}(u_d, u_{\tilde{d}}) \leq d_{\mathcal{GH}}(d, \tilde{d})$	Mikhailov (2025)
ℓ_0 (pair-count; Hamming metric)	Adversarial *sparse* edits of up to k pairs (no magnitude bound); propagation constrained by MST structure.	$\|u_d - u_{\tilde{d}}\|_0 \leq L \, \|d - \tilde{d}\|_0$	This work

Table 1: Stability regimes for the minimax/subdominant ultrametric. Classical uniform bounds control *magnitude* (ℓ_∞/GH), while our Hamming-space (ℓ_0) bound controls the *extent* of change under sparse edits.

1.2 Minmax Ultrametrics in Modern Machine Learning

Zhu et al. (2017) noted that among common hierarchical clustering methods, only single-linkage (the minmax ultrametric) is stable under small perturbations of the input weights (dissimilarities) and consistent in the infinite-sample limit. Specifically, they prove that SLC is the sole method satisfying: *if the number of i.i.d. sample points goes to infinity, the output ultrametric converges (a.s., in Gromov-Hausdorff sense) to the true multiscale structure of the data distribution's support.* Dey et al. (2017) study temporal hierarchical clustering by fitting each time slice with an ultrametric and enforcing small inter-time distortions. They show that the generic ℓ_∞ nearest-ultrametric fit can be unstable under metric perturbations, and therefore replace it with the *minmax ultrametric* u_d, defined as the maximum edge weight along the MST path; u_d is the unique ℓ_∞-closest ultrametric that does not increase any input distance. This choice yields temporally coherent single-linkage dendrograms (since u_d is the single-linkage/minmax ultrametric) while restoring stability in their temporal objective.

Devijver et al. (2024) studied the stability in high-dimensional network inference by inserting a single-linkage hierarchical clustering step before graphical lasso, and proves that the resulting dendrogram, hence the minmax ultrametric underlying single linkage is stable under data perturbations, unlike average linkage. Concretely, the classical two-step decomposition first clusters variables via single linkage on a similarity derived from absolute sample covariances, then fits graphical lasso inside the resulting modules; prior work shows this Step-1 clustering is exactly SLC on that similarity, making the dendrogram the algorithm's structural backbone. The authors provide theoretical bounds controlling distances between dendrograms built from two samples and show in simulations and real data that single linkage based modules are markedly more stable than alternatives, while complete/average linkage can be unstable. Overall, the paper is a recent,

theory-driven application where the minmax ultrametric (via single linkage/MST path-max) is explicitly used to stabilize hierarchical decomposition before sparse graphical model estimation.

Recent methods leverage ultrametrics as algorithmic backbones for broader clustering objectives and pipelines, e.g., showing that center-based objectives can be solved optimally on ultrametrics and producing rich cluster hierarchies (Draganov et al., 2025), and, in density-based settings, building MST/minmax style hierarchies whose slices enjoy formal stability/consistency (Rolle & Scoccola, 2024; Ritzert et al., 2025).

Beyond classical hierarchical clustering, recent work formulates ultrametric fitting as an optimization problem amenable to gradient-based learning and end-to-end training. Learning an ultrametric therefore, amounts to inducing a hierarchy from data rather than merely running a procedural agglomeration. Chierchia & Perret (2019) proposed a continuous optimization framework for learning ultrametrics: they replace the ultrametric constraint by a minmax formulation so one can optimize over ultrametric matrices with standard gradients. Their objective flexibly combines closest ultrametric fidelity with task-driven terms (e.g., Dasgupta's HC objective, cluster-size regularization, triplet constraints), and scales to large graphs with performance comparable to strong agglomerative baselines. In a related vein, Chen et al. (2024) cast tree–Wasserstein regression as ultrametric learning: they learn a tree metric (ultrametric) so that the induced tree–Wasserstein distance approximates the underlying *optimal transport* (OT) distance, using projected gradient descent (projection via a hierarchical map). The learned ultrametric trees outperform several baselines on synthetic and real distributional data. Other optimization-driven approaches include differentiable losses on component trees (Perret & Cousty, 2022) (end-to-end learning of hierarchical segmentations), which directly tune the altitudes (merge levels) of a hierarchy.

Deep methods increasingly enforce ultrametric structure during training. Lapertot et al. (2024) introduce a differentiable ultrametric layer that maps predicted pairwise dissimilarities to an ultrametric, enabling end-to-end learning of hierarchical image segmentations with hierarchy-aware losses (e.g., hierarchical Rand index). In 3D vision (He et al., 2024), ultrametric feature fields impose an ultrametric contrastive loss so latent features satisfy the ultrametric inequality, yielding view-consistent hierarchical segmentations that outperform flat baselines. At a more foundational level, ultrametric neural networks (v-PuNNs) (N'guessan, 2025) with p-adic weights deliver transparent hierarchical representations: each neuron encodes a p-adic ball, guaranteeing a perfectly ultrametric output metric; empirically, these models recover large taxonomies with near-perfect leaf accuracy and zero triangle-inequality violations.

Collectively, these works treat the subdominant ultrametric not merely as a byproduct of single-linkage, but as a stable, computable projection from metrics to trees that supports temporal coherence, statistical stability of dendrograms, and fast algorithmic reductions in modern hierarchical clustering.

Takeaway: The above works show that the minmax (single-linkage) ultrametric is not a relic of classical clustering, but an actively used stability tool in modern hierarchical pipelines: it is chosen precisely because it behaves well under perturbations and admits clean algorithmic structure. Our results refine this picture along a complementary axis: instead of only controlling how *far* an ultrametric can move under metric noise, we control *how many* pairwise relations can change and which parts of the tree they can reach. In settings such as temporal HC or hierarchical graphical models that already rely on u_d for robustness, our Hamming–Lipschitz bounds provide principled tools to (i) localize the effect of sparse perturbations and (ii) identify load-bearing cuts where instability or model misspecification is structurally concentrated.

1.3 Contributions:

At a high level, this paper makes the following contributions.

- **A sparsity-aware view of ultrametric stability.** We move beyond classical ℓ_∞ / Gromov–Hausdorff results and study the minmax ultrametric in a Hamming setting, where the cost of a perturbation is the *number* of pairwise distances that are edited, not how large the edits are.
- **Which pairs can actually change?** We show that sparse edits do *not* propagate arbitrarily through the hierarchy (Theorem 1). Instead, we give a simple structural rule that says exactly which pairs can be affected and when those pairs are guaranteed to remain unchanged.

- **A data-dependent Lipschitz constant.** We introduce a new instance-dependent constant $L_T^\star$ that measures how *fragile* a given hierarchical structure is (Theorem 2). This constant depends only on the tree underlying the ultrametric and gives a bound on how many entries of u_d can move under a sparse perturbation of d.

- **Sharpness and worst-case behavior.** We show that our bound is not only an upper bound but can actually be attained (Theorem 3). In particular, there are examples where changing a *single* pairwise distance forces a quadratic number of ultrametric entries to change, so no substantially smaller, instance-independent bound is possible.

- **When changes almost add up?** For edit sets that *hit* largely disjoint regions of the hierarchy, we prove that the total number of changed entries is very close to the sum of the individual contributions (Corollary 1). In this regime our bound is essentially tight from both sides, giving a clean growth law for the total effect of multiple edits.

- **Simple case studies as diagnostics.** Finally, we use the per-edit score derived from $L_T^\star$ as a diagnostic tool in two small case studies: *(i)* deep embeddings (DINO+UMAP) of CIFAR-10, ImageNet-10, and STL-10, and *(ii)* a superpixel segmentation of the Cameraman image. In both cases, high-score edges line up with empirically fragile parts of the hierarchy, illustrating that our theory can inform practical *vulnerabilitymaps* without aiming at the task.

All proofs of our proposed theorems and corollaries are deferred to Appendix due to space constraint.

2 Notations, Symbols, Assumptions and Preliminaries:

Let $S = \{1, \dots, n\}$ be a finite index set, and let $d : S \times S \to \mathbb{R}_{\geqslant 0}$ be a symmetric dissimilarity with $d(i, i) = 0$ for all $i \in S$. We write $\|\cdot\|$, $\|\cdot\|_\infty$, and $\|\cdot\|_0$ for the ℓ_2 norm, the ℓ_∞ norm, and the ℓ_0 pseudo-norm, respectively (always applied to the vector of upper-triangular entries unless stated otherwise).

We view d in three equivalent ways: (i) as a function on ordered pairs $(i, j) \in S \times S$; (ii) as a symmetric matrix $\mathbf{D} \in \mathbb{R}^{n\times n}$ with entries $\mathbf{D}_{ij} = d(i, j)$; and (iii) as an edge-weight function on the complete undirected graph $G = (V, E)$ with vertex set $V = S$ and edge set $E = \{\{i, j\} : 1 \leqslant i < j \leqslant n\}$. When convenient we write $d(e)$ for the weight of an edge $e = \{i, j\} \in E$.

Let $T = (V, E(T))$ denote the (tie-broken) minimum spanning tree (MST) of d on G. For each tree edge $e = \{a, b\} \in E(T)$, we write $C_e = (A_e, B_e)$ for the fundamental cut obtained by removing e from T, so that $A_e, B_e \subseteq V$ are the vertex sets of the two connected components of $T \setminus \{e\}$. We abbreviate $w_e := d(e)$ and define the *alternative cut minimum*

$$w_e^+(d) \;:=\; \min\big\{\, d(x, y) : x \in A_e,\; y \in B_e,\; \{x, y\} \neq e \,\big\}, \tag{2}$$

and the corresponding *cut gap*

$$\Delta_e(d) \;:=\; w_e^+(d) - w_e \;>\; 0, \tag{3}$$

which is strictly positive under our tie-breaking convention.

Given a perturbed dissimilarity $\tilde{d} : S \times S \to \mathbb{R}_{\geqslant 0}$, we measure the size of the perturbation by the number of edited pairs, i.e., in the Hamming/ℓ_0 sense:

$$\|d - \tilde{d}\|_0 \;:=\; \#\big\{\, \{i, j\} \subseteq S : i < j,\; d(i, j) \neq \tilde{d}(i, j) \,\big\}. \tag{4}$$

Let $(h_r)_{r=1}^{n-1}$ denote the increasing list of MST edge weights $h_r \in \{w_e : e \in E(T)\}$, and similarly $(\tilde{h}_r)_{r=1}^{n-1}$ for the MST of $\tilde{d}$.

Ultrametric operator in index notation. In the introduction we viewed the subdominant (min–max) ultrametric as the output of an operator

$$U : (X, d) \mapsto (X, u_d). \tag{5}$$

In this finite-index setting we work directly with the indexed version: for each dissimilarity d on S, we denote by $u_d : S \times S \to \mathbb{R}_{\geqslant 0}$ the associated subdominant ultrametric, and we write

$$U(d) = u_d, \qquad u_d(i, j) \text{ for the } (i, j)\text{-entry of the ultrametric.} \tag{6}$$

All of our subsequent theorems and inequalities will be stated in terms of indices $i, j \in S$, the MST T on S, and the entries of d and u_d.

To ensure a unique MST, we adopt the standard assumption of lexicographic tie-breaking: whenever two edges have identical weights, their order is resolved by a fixed, secondary lexicographic ordering (for instance, based on endpoint indices or edge identifiers). This convention is not restrictive, it is a widely accepted device in both theoretical analyses and algorithmic implementations of MST-related problems. Recent works across theory and systems routinely employ the same assumption to guarantee determinism and analytical clarity, including sublinear and dynamic formulations (Patlin & van den Brand, 2025; de Vos & Grilnberger, 2025), massively parallel and distributed MST algorithms (Azarmehr et al., 2025; Sanders & Schimek, 2023), and polymatroid-based theoretical generalizations (Harb et al., 2023). Thus, lexicographic tie-breaking serves as a benign technical convention rather than a substantive limitation, ensuring well-definedness without affecting optimality or generality.

Assumption 1. *Fix a deterministic total order $\prec$ on edges (e.g., lexicographic). We compare edges by the lexicographic pair $(w(e), \mathrm{rank}_{\prec}(e))$. Equivalently, conceptually perturb*

$$w'(e) \;=\; w(e) \;+\; \eta\, \mathrm{rank}_{\prec}(e), \qquad \textit{with a single infinitesimal } \eta > 0, \tag{7}$$

and run Kruskal on w' (ties in w are broken by $\prec$). This yields a unique MST T and a strictly ordered MST edge list.

Under this convention, Kruskal adds all within component edges before the cut edge; the next lemma records this.

Lemma 1. *Let $e_t \in E(T)$ with weight w_{e_t} and fundamental cut $C_{e_t} = (A_{e_t}, B_{e_t})$. Then every MST edge whose endpoints both lie in A_{e_t} (resp. in B_{e_t}) has weight* strictly *less than w_{e_t}.*

Proof. The proof is trivial. Let $e_t \in E(T)$ with weight w_{e_t} and fundamental cut $C_{e_t} = (A_{e_t}, B_{e_t})$. Then every MST edge whose endpoints both lie in A_{e_t} (resp. in B_{e_t}) has weight *strictly* less than w_{e_t}.

□

We now formally define the minmax ultrametric below:

Definition 1 (Subdominant (minmax) ultrametric)**.** *Given a finite set S with dissimilarity function $d : S \times S \to \mathbb{R}_{\geqslant 0}$, the minmax subdominant ultrametric $u_d : S \times S \to \mathbb{R}_{\geqslant 0}$ is the largest ultrametric dominated by d, equivalently*

$$u_d(i, j) \;=\; \min_{P \in \mathcal{P}(i,j)} \; \max_{(i,j) \in P} d(i, j) \quad (i \neq j), \tag{8}$$

where $\mathcal{P}(i, j)$ is the set of all paths that connect from i to j, and P is a specific path in that set. If $e = \{i, j\}$ denotes the edge connecting i and j, then we can write $u_d(i, j) = \min_{P \in \mathcal{P}(i,j)} \max_{e \in P} d(e)$.

Definition 2 (Path and bottleneck)**.** *For $i, j \in S$, let $\mathcal{P}(i, j)$ denote the set of all finite paths from i to j. A path $P = (v_0, v_1, \ldots, v_m)$ is a finite sequence with $v_0 = i$, $v_m = j$, and $v_{t-1} \neq v_t$ for $t \in \{1, \ldots, m\}$. The* bottleneck *of P is defined as $M(P) := \max_{1 \leqslant t \leqslant m} d(v_{t-1}, v_t)$.*

The bottleneck captures the largest edge weight along a given path, and forms the basis for defining the minmax ultrametric transform.

Definition 3 (Path bottleneck definition of minmax ultrametric)**.** *The minmax ultrametric transform of a dissimilarity d is the map $u_d : S \times S \to \mathbb{R}_{\geqslant 0}$ defined by $u_d(i,j) := \min_{P \in \mathcal{P}(i,j)} M(P)$, for $i \neq j$, with $u_d(i,i) = 0$. Equivalently, $u_d(i,j) = \min_{P \in \mathcal{P}(i,j)} \max_{(i,j) \in P} d(i,j)$.*

2.1 Important Definitions and Lemmas

We recall two MST facts that underpin our analysis. First, the MST characterization identifies the iVAT (subdominant ultrametric) distance between i and j with the heaviest edge on their path in an MST, so visual structure reduces to path bottlenecks. Second, the cut property with uniqueness ensures that a strictly lightest cut edge appears in every MST, giving a natural notion of a gap across the cut. In later sections, we use these lemmas to define cut-gap certificates, bound edge-order inversions, and localize exactly which entries of the reordered matrix can change under perturbations. These are common Lemmas on MST, one can easily find them in the standard algorithms textbook by Cormen et al.

Lemma 2 (MST characterization)**.** *If T is any minimum spanning tree (MST) of the complete graph with weights given by the dissimilarity function d, then $u_d(i,j) = \max_{e \in \mathrm{path}_T(i,j)} d(e) \quad \forall i \neq j$.*

Lemma 3 (Cut property, with uniqueness)**.** *Let (A, B) be any cut of V and let $e^\star \in E$ be an edge with one endpoint in A and one in B. If $d(e^\star) < d(f)$ for every other cut edge f across (A, B), then $e^\star$ belongs to* every *MST of d.*

3 Sparsity-Localized Perturbations and Pairwise Impact

Our first theorem establishes a localization and upper bound principle for the minmax ultrametric under sparse edits. At the level of the MST, it shows that if the tree path between two points does not interact with a small set of *affected* cuts, then their ultrametric distance can only decrease, and under a simple condition, it in fact remains unchanged. More globally, the theorem upper-bounds the total number of pairs that can change in the ultrametric space by a *union of cut rectangles* associated with a small collection of tree edges, giving an explicit combinatorial control on the support of $u_d - u_{\tilde{d}}$. For single-linkage clustering, this means that the dendrogram is locally stable: only branches attached to these affected cuts can move, and the rest of the hierarchy is rigid. Thus, Theorem 1 provides both a structural description of where changes may occur and a worst-case bound on how many pairwise merge heights can be altered by a sparse perturbation.

Theorem 1 (Localization of ultrametric under sparse edge edits)**.** *Let $d : \binom{V}{2} \mapsto \mathbb{R}$, and let $T = (V, E(T))$ be the (tie-broken under Assumption 1) MST of d. For $e = \{a, b\} \in E(T)$ let $C_e = (A_e, B_e)$ be its fundamental cut in T, and write $w_e := d(e)$. Let $F \subseteq \binom{V}{2}$ be a set of edited edges and let $\tilde{d}$ be any dissimilarity with $\tilde{d}(e) = d(e)$ for all $e \notin F$ (no restriction on $e \in F$).*

For $i \neq j$ let $P_T(i,j)$ be the unique i–j path in T. From MST bottleneck representation $u_d(i,j) = \max_{e \in P_T(i,j)} w_e$, the following holds:

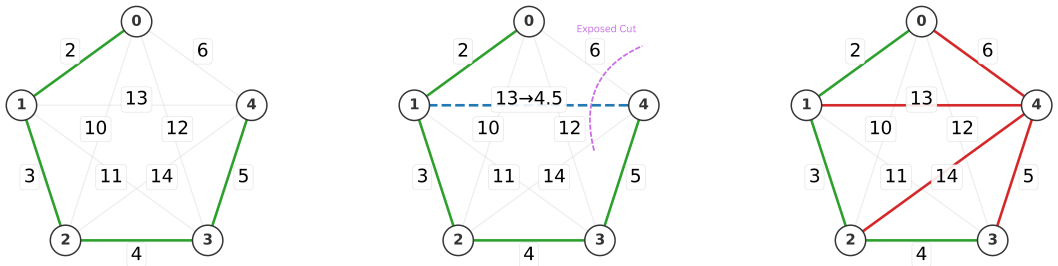
(i) (Monotone upper bound, no edited T-edges on the path) If $P_T(i,j) \cap F = \varnothing$, then $u_{\tilde{d}}(i,j) \leq u_d(i,j)$.

(ii) (Sufficient conditions for equality) If $P_T(i,j) \cap F = \varnothing$ and for every $e \in P_T(i,j)$, all edited edges $f \in F$ crossing C_e satisfy $\tilde{d}(f) \geqslant w_e$, then $u_{\tilde{d}}(i,j) = u_d(i,j)$.

(iii) (Pair-count bound for possible changes) Define the set of potentially affected MST edges $\mathcal{E} := \{\, e \in E(T) \mid e \in F \text{ or } \exists f \in F \text{ crossing } C_e,\ \tilde{d}(f) < w_e \,\}$. Then the number of nodes with changed ultrametric distance will be upper bound as follows:

$$|\{(i,j) : u_{\tilde{d}}(i,j) \neq u_d(i,j)\}| = \|u_d - u_{\tilde{d}}\|_0 \leqslant \left| \bigcup_{e \in \mathcal{E}} (A_e \times B_e) \right| = \frac{1}{2} n(n-1) - \frac{1}{2} \sum_{t=1}^{m} |C_t| \, (|C_t| - 1), \tag{9}$$

where $C_1, \ldots, C_m$ are the vertex sets of the connected components of the forest $T - \mathcal{E}$.



(a) (b) (c)

Figure 1: Illustration of Theorems 1 and 2. (a) Initial MST (green). (b) A sparse off-tree edit on edge $\{1, 4\}$ (dotted-blue) that exposes the MST edge $\{4, 5\}$. (c) The affected region (red), given by the cut rectangle $\{0, 1, 2, 3\} \times \{4\}$, upper-bounds the total change in the ultrametric space, i.e $\|u_d - u_{\tilde{d}}\|_0$.

Theorem 1 gives a first, coarse localization result: it tells us that all changes in the ultrametric are confined to a union of fundamental rectangles $A_e \times B_e$ over a small set of MST edges $\mathcal{E}$ determined by the edit pattern. However, this description is still somewhat entangled across edits and does not yet look like a Lipschitz inequality in the Hamming sense: it upper-bounds $\|u_d - u_{\tilde{d}}\|_0$ in terms of a global "affected edge" set, rather than decomposing the effect of each edited pair.

To move toward a true Hamming–Lipschitz bound of the form $\|u_d - u_{\tilde{d}}\|_0 \leq L_T^\star \, \|d - \tilde{d}\|_0$, we refine the picture at the level of a single edited edge. For each edited pair f, we isolate the subset of tree edges whose cuts are actually made cheaper by f; these are the *exposed cuts* $\Xi(f)$. The next theorem shows that the pairs whose ultrametric values can change due to f lie inside the union of rectangles attached to $\Xi(f)$, and that summing these per-edit regions leads to an instance-dependent Lipschitz constant $L_T^\star$ governing how many ultrametric entries can change under an arbitrary sparse perturbation.

Theorem 2 (Hamming–Lipschitz bound via exposed cuts)**.** *Let $d : \binom{V}{2} \to \mathbb{R}_{\geqslant 0}$ be a dissimilarity on a finite set V, and let $T = (V, E(T))$ be the (tie-broken under Assumption 1) MST of d. For each tree-edge $e = \{a, b\} \in E(T)$, let $C_e = (A_e, B_e)$ be its fundamental cut in T, and write $w_e := d(e)$.*

Let $F \subseteq \binom{V}{2}$ be an arbitrary set of edited pairs, and let $\tilde{d}$ be any dissimilarity such that $\tilde{d}(e) = d(e)$ for all $e \notin F$ (no restriction on $e \in F$). For an edited pair $f = \{x, y\} \in F$, define the set of exposed cuts *by*

$$\Xi(f) \;:=\; (\{f\} \cap E(T)) \;\cup\; \big\{\, e \in E(T) : \; f \text{ crosses } C_e \;\text{ and }\; \tilde{d}(f) < w_e \,\big\}. \tag{10}$$

Then the set of unordered pairs whose ultrametric values change is contained in the union of the fundamental rectangles attached to the exposed cuts:

$$\mathcal{C} \;:=\; \big\{\{i, j\} \in \tbinom{V}{2} : u_{\tilde{d}}(i, j) \neq u_d(i, j)\big\} \;\subseteq\; \bigcup_{f \in F} \; \bigcup_{e \in \Xi(f)} (A_e \times B_e). \tag{11}$$

Consequently,

$$\|u_d - u_{\tilde{d}}\|_0 \;=\; |\mathcal{C}| \;\leqslant\; \left| \bigcup_{f \in F} \; \bigcup_{e \in \Xi(f)} (A_e \times B_e) \right| \tag{12}$$

In particular, defining the per-edit affected size

$$S_{\text{union}}(f) \;:=\; \left| \bigcup_{e \in \Xi(f)} (A_e \times B_e) \right|, \qquad L_T^\star \;:=\; \max_{f \in \binom{V}{2}} S_{\text{union}}(f), \tag{13}$$

we obtain the instance-dependent Hamming–Lipschitz inequality

$$\|u_d - u_{\tilde{d}}\|_0 \;\leqslant\; \sum_{f \in F} S_{\text{union}}(f) \;\leqslant\; L_T^\star \, |F| \;= L_T^\star \|d - \tilde{d}\|_0 \leqslant \binom{n}{2}. \tag{14}$$

We now interpret the ℓ_0 upper bound from Theorem 2, we next ask whether its dependence on the exposed-cut union and the instance parameter $L_T^\star$ is merely a proof artifact or an intrinsic feature of the ultrametric operator. Our next theorem, Theorem 3, provides this calibration. It builds single-edit perturbations that achieve the Theorem 2 bound exactly, both when the edit is on the tree and when it is off tree, thereby showing that our union of rectangles control is tight at the level of individual edits. It also exhibits worst case metrics where a single off tree edit flips a quadratic number of pairs, proving that no universal, subquadratic Lipschitz constant can replace the instance-dependent one.

Theorem 3 (Analysis of the Hamming–Lipschitz bound)**.** *Let $d : \binom{V}{2} \to \mathbb{R}_{\geqslant 0}$ be a dissimilarity and let $T = (V, E(T))$ be the (tie-broken under Assumption 1) MST of d. For $e \in E(T)$ write its fundamental cut $C_e = (A_e, B_e)$ and $w_e = d(e)$. For an edited pair $f = \{x, y\}$ and edited dissimilarity $\tilde{d}$ with $\tilde{d}(g) = d(g)$ for $g \notin \{f\}$, define*

$$\Xi(f) \;:=\; (\{f\} \cap E(T)) \,\cup\, \{\, e \in E(T) :\; f \text{ crosses } C_e \text{ and } \tilde{d}(f) < w_e \,\}, \qquad S_{\mathrm{union}}(f) \;:=\; \Big|\, \bigcup_{e \in \Xi(f)} (A_e \times B_e) \,\Big|. \tag{15}$$

Then:

(i) (Tree-edge edit) *If $f = e \in E(T)$, there exists $\tilde{d}$ supported on $\{f\}$ such that*

$$\|u_d - u_{\tilde{d}}\|_0 \;=\; |A_e|\,|B_e| \;=\; S_{\mathrm{union}}(f). \tag{16}$$

(ii) (Off-tree edit) *If $f \notin E(T)$, there exist weights (achieved by an arbitrarily small tie-breaking perturbation of d along $P_T(x, y)$) and a choice of $\tilde{d}$ supported on $\{f\}$ with $\tilde{d}(f) < \min_{e \in P_T(x,y)} w_e$ such that*

$$\|u_d - u_{\tilde{d}}\|_0 \;=\; \Big|\, \bigcup_{e \in \Xi(f)} (A_e \times B_e) \,\Big| \;=\; S_{\mathrm{union}}(f). \tag{17}$$

(iii) (Necessity of instance dependence) *There exist $d, \tilde{d}$ with $\|d - \tilde{d}\|_0 = 1$ such that*

$$\|u_d - u_{\tilde{d}}\|_0 \;=\; \Theta(n^2). \tag{18}$$

Consequently, no universal subquadratic $c(n) = o(n^2)$ can satisfy $\|u_d - u_{\tilde{d}}\|_0 \leqslant c(n)\,\|d - \tilde{d}\|_0$ on all instances; the instance-dependent constant is necessary.

Theorem 3 shows that the structure revealed by our upper bound is not a proof artifact: for a single edit on tree, or off tree, the changed pairs coincide exactly with the union of exposed fundamental cut rectangles, and in the worst case a single off tree edit flips a quadratic number of pairs, proving that any universal subquadratic constant is impossible and that instance dependence is necessary. This calibration raises the practical question for multiple edits: when does the union bound from Theorem 2 become an accurate predictor rather than a loose ceiling? Corollary 1 answers this by formalizing a minimal overlap regime in which per edit affected regions intersect negligibly; in this setting the total number of changes is within a vanishing factor of the sum of the per edit region sizes, yielding a two-sided (near-additive) characterization. Intuitively, this pins down the operational conditions, dispersed edits that expose mostly distinct cuts under which our instance dependent pair count bound is tight, while Theorem 3 explains why no stronger, overlap agnostic constant can exist.

Corollary 1. *Fix a minimum spanning tree T of d. Let $F \subseteq \binom{V}{2}$ be a finite edit set such that the per-edit affected regions*

$$\mathcal{R}(f) \;:=\; \bigcup_{e \in \Xi(f)} (A_e \times B_e) \tag{19}$$

have negligible pairwise overlaps:

$$|\mathcal{R}(f) \cap \mathcal{R}(f')| \;=\; o(S_{\mathrm{union}}(f)) \quad \text{for all } f \neq f', \qquad S_{\mathrm{union}}(f) := |\mathcal{R}(f)|. \tag{20}$$

Then there exist edited weights $\tilde{d}$ (via infinitesimal tie-breaking consistent with T) such that

$$(1 - o(1)) \sum_{f \in F} S_{\text{union}}(f) \;\leqslant\; \|u_d - u_{\tilde{d}}\|_0 \;\leqslant\; \sum_{f \in F} S_{\text{union}}(f). \tag{21}$$

Hence, in the minimal-overlap regime, the Hamming–Lipschitz inequality, $\|u_d - u_{\tilde{d}}\|_0 \leqslant L_T^\star |F|$ is order-tight*: the realized change in the ultrametric lies within a vanishing factor of the upper bound, yielding a two-sided (sandwich) characterization of its asymptotic growth.*

4 Empirical Case Studies

Our primary contribution is theoretical. The preceding sections develop a Hamming-Lipschitz stability theory for the subdominant ultrametric, characterize its instance-dependent Lipschitz constant $L_T^\star$, and construct worst-case examples where a single edit flips $\Theta(n^2)$ ultrametric entries. We now present two small case studies whose sole purpose is to illustrate how these quantities behave in concrete settings and how they can be used to construct *vulnerability maps* of hierarchical structures. We do *not* aim to beat task-specific state-of-the-art methods in segmentation or representation learning; instead we show that the same combinatorial objects that appear in our proofs lead to interpretable diagnostics in practice.

4.1 Experimental protocol

In both case studies, we work with a metric space (V, d) and its (tie-broken) minimum spanning tree (MST) $T = (V, E(T))$. We compute the subdominant ultrametric u_d via the MST bottleneck representation, and we use symbolic tie-breaking as in Assumption 1 so that the MST and its edge order are unique without numerically perturbing the weights.

For each tree edge $e \in E(T)$, removing e splits the tree into components A_e and B_e. We define the tree-edge risk score $S_{\text{union}}(e) = |A_e|\,|B_e|$, which is exactly the cardinality of the fundamental rectangle $A_e \times B_e$ that appears in the union-of-rectangles bound in Eq. 16. For a general edited pair f, Theorem 2 associates a region, $R(f) = \bigcup_{e \in \Xi(f)} A_e \times B_e$, where $\Xi(f)$ is the set of exposed cuts, and proves that

$$\left\|u_d - u_{\tilde{d}}\right\|_0 \;\leqslant\; \sum_{f \in F} |R(f)| \;=\; \sum_{f \in F} S_{\text{union}}(f) \;\leqslant\; L_T^\star\, |F|. \tag{22}$$

In our experiments, we instantiate this inequality in two complementary ways:

- **Sparse adversarial edits on representation graphs (Case study 4.2).** Given a budget of m edited edges, we compare a *structural attack* that targets the m edges with largest $S_{\text{union}}(e)$ against a *random baseline* that edits m edges uniformly at random from the same pool, and measure the resulting Hamming distortion of u_d.
- **Single-edge and safe-edit analysis in image segmentation (Case study 4.3).** On a super-pixel graph for the Cameraman image, we take $F = \{e\}$ and study the realized change in the induced segmentation when the weight of a single edge e is strongly decreased or increased, then aggregate these per-edge impacts into *safe-edit* curves that show how much damage can occur when we are allowed to edit only edges below a given risk threshold.

4.2 Case study 1: Vulnerability maps of deep embeddings

We first consider high-dimensional embeddings derived from deep networks. We use three 10 classes image datasets, namely CIFAR-10, ImageNet-10 and STL-10. For each dataset we start from precomputed DINO-ViT features, followed by UMAP to 20 dimensions, and we subsample a few thousand points from the test set for efficiency. On each embedding, we build a complete Euclidean graph and compute its tie-breaking MST and induced ultrametric u_d.

Instance-dependent risk profiles. For each MST edge we compute $S_{\text{union}}(e)$ and plot the histogram of $\log_{10} S_{\text{union}}(e)$. Across all three datasets, the distribution is highly skewed: a large spike at low values (many leaf edges whose cuts isolate very small components) and a long, thin tail at high values (a small number of *bridges* whose cuts produce two large components). These histograms can be viewed as empirical vulnerability profiles: they visualize how much mass of the instance-dependent Lipschitz constant $L_T^\star$ is concentrated in a small set of structurally dangerous edges.

Adversarial versus random sparse edits. To probe the Hamming–Lipschitz theory directly, we simulate sparse perturbations of the metric. Fix a small budget m (we have used 10% of the MST edges). We consider two editing strategies:

- **Structural attack (top-risk):** select the top-m tree edges with largest $S_{\text{union}}(e)$.
- **Random baseline:** select m tree edges uniformly at random.

In both cases, we construct a perturbed metric $\tilde{d}$ by sharply inflating the weight of each selected edge so that it becomes much heavier than any other edge crossing the corresponding cut, mimicking the single-edit constructions in Theorem 3. We then recompute the MST and the minmax ultrametric $u_{\tilde{d}}$ and measure the normalized Hamming distance $\Delta_0(u_d, u_{\tilde{d}}) = \frac{\|u_d - u_{\tilde{d}}\|_0}{\binom{n}{2}}$.

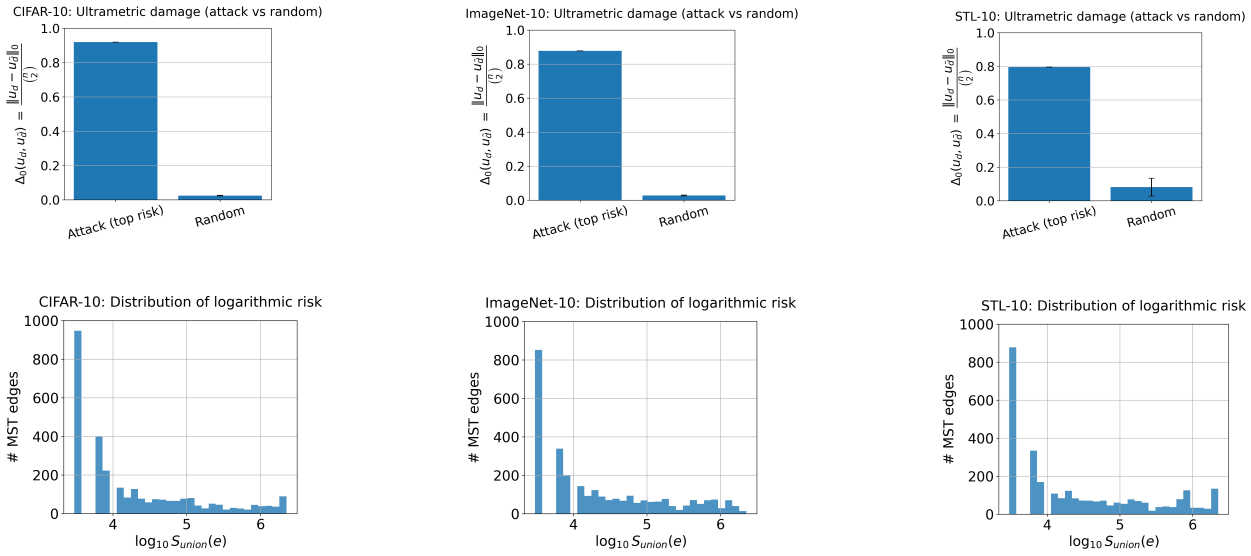


Figure 2: **Top row:** Ultrametric damage $\Delta_0(u_d, u_{\tilde{d}})$ under targeted (top-risk) vs. random edge perturbations for CIFAR-10, ImageNet-10, and STL-10. **Bottom row:** Distributions of structural risk scores $\log_{10} S_{\text{union}}(e)$ over MST edges for the same three datasets.

Results. For all three datasets, we find that:

- From Fig. 2 (top row), we can observe that the top-risk attack consistently causes orders-of-magnitude larger ultrametric damage than random edits with the same budget. Even at very small budgets where the random baseline leaves $\Delta_0(u_d, u_{\tilde{d}})$ essentially zero, targeting the top-S_{union} edges can already flip a significant fraction of ultrametric entries.
- From Fig. 2, the bottom row shows a common qualitative pattern across all three datasets: a large spike at low values of $S_{\text{union}}(e)$ (many low-risk, leaf-like edges) and a long, thin tail at higher values (a much smaller set of potentially dangerous *bridge* edges). The histograms show that every dataset

concentrates most of its Lipschitz mass in a tiny high-risk tail, and the bar plots (top row) confirm that editing exactly those edges is what drives the large ultrametric distortion.

Taken together, these experiments provide an empirical counterpart to our theoretical picture: S_{union} identifies a small set of $load - bearing$ edges where Theorem 3's quadratic worst case nearly materializes, while most MST edges are structurally safe. Corollary 1 explains when the union-of-rectangles upper bound becomes nearly additive; the histograms and adversarial versus random experiments show that this regime is meaningful for real representation graphs.

4.3 Case study 2: Fragility maps for a single image segmentation

We next turn to a controlled, low-dimensional example based on the classical Cameraman image. We convert the grayscale image to a float in $[0, 1]$ and oversegment it into K *simple linear iterative clustering* (SLIC) superpixels. Each superpixel becomes a node in a region adjacency graph; we connect superpixels whose labels touch in the pixel grid and equip each edge with a simple feature distance combining mean intensity and normalized centroid coordinates. We then compute the tie-broken MST T of this graph and its induced ultrametric u_d.

Figure 3: The classic Cameraman image.

From T we obtain a base segmentation by cutting the $K - 1$ heaviest tree edges and taking the resulting connected components. This is the standard interpretation of single linkage as cutting an MST at a fixed level. We use this segmentation as our *reference* partition and map it back to pixels by assigning each superpixel its component label.

Per-edge fragility. We first obtain a reference segmentation by running single linkage on the superpixel MST and cutting at a fixed number of clusters K; this partition is mapped back to pixels and used as the baseline segmentation. For each tree edge $e = \{a, b\}$, we then compute:

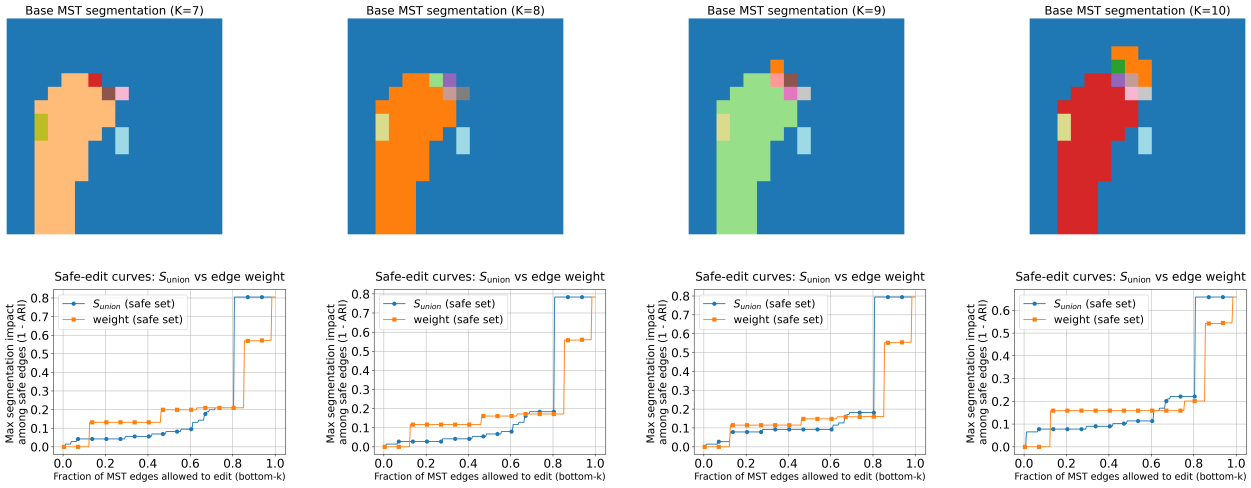
- its structural risk $S_{\text{union}}(e) = |A_e|\,|B_e|$;
- its naive weight $w(e)$ (local boundary contrast);
- and its *worst-case segmentation impact*: we construct two perturbed graphs in which the weight of e is multiplied by a small factor (10^{-2}) or a large factor (10^{2}), recompute the MST and the induced segmentation for each perturbation, and set

$$\text{impact}(e) \;=\; \max\Big\{1 - \text{ARI}\big(\text{base}, \text{decrease}(e)\big),\; 1 - \text{ARI}\big(\text{base}, \text{increase}(e)\big)\Big\}, \tag{23}$$

  where ARI is the adjusted Rand index between superpixel partitions.

Safe-edit experiment. For each MST edge e we thus have its structural score $S_{\text{union}}(e) = |A_e||B_e|$, its local weight $w(e)$, and its precomputed worst-case impact impact(e). For each scoring rule (S_{union} or w), we rank MST edges in *ascending* order (small score = "safe"). For a given prefix size k (shown on the x-axis as a fraction of all tree edges), we treat the bottom-k edges under that score as the *safe-to-edit* set and take the maximum of impact(e) over this set. Plotting this worst-case impact as a function of k yields two safe-edit curves: one when "safe" edges are chosen by S_{union}, and one when they are chosen by raw weights.

In our experiments, the curve based on S_{union} consistently attains lower maximal damage for a given safe budget, indicating that low-score edges under the structural metric can be edited with far less impact than edges deemed "light" by their standalone weights. Thus, even in this simple vision setting, the tree-aware score derived from our Hamming–Lipschitz analysis provides a more reliable notion of safe directions to modify the data than a purely local boundary-contrast heuristic.



(a) $K = 7$ Curves (b) $K = 8$ Curves (c) $K = 9$ Curves (d) $K = 10$ Curves

Figure 4: Safe-edit curves (bottom row) and corresponding reference segmentations (top row) for Cameraman image across $K = 7$–10. For each K, we report worst-case $1 - \mathrm{ARI}$ when edits are restricted to bottom-k MST edges ranked by $S_{\mathrm{union}}(e)$ or by $w(e)$.

5 Conclusion

We developed a sparsity-aware stability theory for the subdominant (minmax) ultrametric, complementing classical ℓ_∞/Gromov–Hausdorff results with a Hamming–Lipschitz perspective that controls the *extent* of change under sparse edits. Our analysis shows that propagation is mediated by the MST: only pairs whose tree paths traverse edited or newly exposed cuts can change, yielding a union-of-rectangles description and an instance-dependent constant $L_T^\star$. We further proved that this instance dependence is unavoidable by exhibiting single-edit constructions with $\Theta(n^2)$ changes, and identified regimes where our upper bound is order-tight. Together, these results offer a structural, interpretable account of how local perturbations can (or cannot) ripple through single-linkage hierarchies, and motivate using the associated risk scores as practical diagnostics for locating robust versus load-bearing edges in real data.

References

Amir Azarmehr, Soheil Behnezhad, Rajesh Jayaram, Jakub Łącki, Vahab Mirrokni, and Peilin Zhong. Massively parallel minimum spanning tree in general metric spaces. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 143–174. SIAM, 2025.

Gunnar E Carlsson, Facundo Mémoli, et al. Characterization, stability and convergence of hierarchical clustering methods. *J. Mach. Learn. Res.*, 11(Apr):1425–1470, 2010.

Samantha Chen, Puoya Tabaghi, and Yusu Wang. Learning ultrametric trees for optimal transport regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 20657–20665, 2024.

Giovanni Chierchia and Benjamin Perret. Ultrametric fitting by gradient descent. *Advances in neural information processing systems*, 32, 2019.

Samir Chowdhury, Facundo Mémoli, and Zane T Smith. Improved error bounds for tree representations of metric spaces. *Advances in Neural Information Processing Systems*, 29, 2016.

TH Cormen, CE Leiserson, RL Rivest, and C Stein. Mit press & mcgraw-hill; 2001. *Introduction to Algorithms, 2nd edition.[Google Scholar]*.

Tijn de Vos and Mara Grilnberger. Dynamic matroids: Base packing and covering. *arXiv preprint arXiv:2511.15460*, 2025.

Emilie Devijver, Rémi Molinier, and Mélina Gallopin. Stable network inference in high-dimensional graphical model using single-linkage. *arXiv preprint arXiv:2406.09865*, 2024.

Tamal K. Dey, Alfred Rossi, and Anastasios Sidiropoulos. Temporal Hierarchical Clustering. In Yoshio Okamoto and Takeshi Tokuyama (eds.), *28th International Symposium on Algorithms and Computation (ISAAC 2017)*, volume 92 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 28:1–28:12, Dagstuhl, Germany, 2017. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-054-5. doi: 10.4230/LIPIcs.ISAAC.2017.28. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ISAAC.2017.28.

Andrew Draganov, Pascal Weber, Rasmus Skibdahl Melanchton Jørgensen, Anna Beer, Claudia Plant, and Ira Assent. I want'em all (at once)–ultrametric cluster hierarchies. *arXiv preprint arXiv:2502.14018*, 2025.

Elfarouk Harb, Kent Quanrud, and Chandra Chekuri. Convergence to lexicographically optimal base in a (contra) polymatroid and applications to densest subgraph and tree packing. *arXiv preprint arXiv:2305.02987*, 2023.

John A Hartigan. Statistical theory in clustering. *Journal of classification*, 2(1):63–76, 1985.

Haodi He, Colton Stearns, Adam W Harley, and Leonidas J Guibas. View-consistent hierarchical 3d segmentation using ultrametric feature fields. In *European Conference on Computer Vision*, pp. 268–286. Springer, 2024.

Anil K Jain and Richard C Dubes. *Algorithms for clustering data*. Prentice-Hall, Inc., 1988.

Raphael Lapertot, Giovanni Chierchia, and Benjamin Perret. End-to-end ultrametric learning for hierarchical segmentation. In *International Conference on Discrete Geometry and Mathematical Morphology*, pp. 286–297. Springer, 2024.

A Martínez-Pérez. Gromov–hausdorff stability of linkage-based hierarchical clustering methods. *Advances in Mathematics*, 279:234–262, 2015.

IN Mikhailov. Ultrametric spaces and clouds. *arXiv preprint arXiv:2501.19346*, 2025.

Gnankan Landry Regis N'guessan. v-punns: van der put neural networks for transparent ultrametric representation learning. *arXiv preprint arXiv:2508.01010*, 2025.

Gryphon Patlin and Jan van den Brand. Sublinear-time algorithm for mst-weight revisited. In *2025 Symposium on Simplicity in Algorithms (SOSA)*, pp. 46–53. SIAM, 2025.

Benjamin Perret and Jean Cousty. Component tree loss function: Definition and optimization. In *International Conference on Discrete Geometry and Mathematical Morphology*, pp. 248–260. Springer, 2022.

Martin Ritzert, Polina Turishcheva, Laura Hansel, Paul Wollenhaupt, Marissa A Weis, and Alexander S Ecker. Hierarchical clustering with maximum density paths and mixture models. *arXiv preprint arXiv:2503.15582*, 2025.

Alexander Rolle and Luis Scoccola. Stable and consistent density-based clustering via multiparameter persistence. *Journal of Machine Learning Research*, 25(258):1–74, 2024.

Peter Sanders and Matthias Schimek. Engineering massively parallel mst algorithms. In *2023 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pp. 691–701. IEEE, 2023.

Roger N Shepard. The analysis of proximities: multidimensional scaling with an unknown distance function. i. *Psychometrika*, 27(2):125–140, 1962.

Robin Sibson. *Mathematical taxonomy*. Wiley, 1971.

Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301):236–244, 1963.

Dekang Zhu, Dan P Guralnik, Xuezhi Wang, Xiang Li, and Bill Moran. Statistical properties of the single linkage hierarchical clustering estimator. *Journal of Statistical Planning and Inference*, 185:15–28, 2017.